nature

www.nature.com/nature

Vol 452 | Issue no. 7188 | 10 April 2008

The path to productive partnerships

Research funders and institutions do too little to sustain the collaborations that they encourage.

A rich culture of collaboration is one of the valued treasures of the scientific community, and deservedly so. It is only natural that universities and funding agencies would want to encourage this approach. It is a shame that their support is only skin-deep.

Many funders have designed programmes that specifically encourage interdisciplinary or international collaborations. Such programmes can enhance funding for collaborations that do not fit easily into traditional grant programmes. More worrying are the initiatives that attempt social engineering, demanding that scientists include researchers from, say, a particular region in their funding application. As a result, efforts are wasted in collaborations that are motivated more by the funds themselves than by scientific necessity.

But most concerning of all is the fact that merely funding a culture of collaboration does too little to make it robust. More support by agencies, and foresight by researchers themselves, needs to be devoted to anticipating and coping with the traps that can lie ahead.

Financial incentives will not teach collaborators how to work as a team. Interdisciplinary endeavours by definition bring together researchers from diverse scientific cultures, with respective languages and priorities. Computer scientists, for instance, may think nothing of making their data freely available on the Internet. But in highly competitive biomedical fields, in which whispers heard at meetings can quickly be turned into papers bearing the eavesdroppers' names, sharing too soon can have unintended consequences.

Meanwhile, crossing political borders carries the risk of complications that can befuddle expert regulators, let alone inexperienced researchers. The rules regarding ownership of intellectual property vary from country to country, as do ethical regulations on experiments involving humans or animals. And when cases of misconduct arise, it is unclear which country or regulatory authority has jurisdiction (see page 686).

These difficulties are not limited to large-scale, high-budget projects, yet researchers too often enter into the complex landscape of collaboration with little, if any, thought about the complications that can arise — even from a small-scale alliance (for examples, see page 682). The culture of informal collaboration feeds scientific innovation and productivity, but it must be balanced by good planning. Wise collaborators will map out their endeavours and responsibilities early in the relationship, keeping a written record of their agreement. These plans need not be limiting: they can and should be amended as the project progresses.

"Researchers too often enter into collaborations with little thought about the complications that can arise."

Meanwhile, an early conversation about data ownership and authorship can prevent the hostility that may arise when such issues are disputed later in the relationship. This discussion should include, from day one, any proposed industrial contacts and contracts. When publication of the results is the goal, reagents, methods and data used and developed during the collaboration need to be kept accessible to the researchers, even if licensed to a second party. Without a written agreement, access to these reagents cannot be ensured.

But expecting researchers themselves to invent such principles from scratch is a recipe for wasted effort. What is more, when researchers fall out, the burden of clearing up the mess can fall heavily on their institutions. Thus the institutions, and funding agencies too, should support collaborations not only with facilities and funds but also through logistical support in the form of guidelines, checklists, model agreements and training — or even counselling, if it should come to that. In that way they can help researchers and themselves avoid the pain, and benefit from the pleasures and insights that collaboration should bring. ■

Defining 'natural'

Visceral reactions to an act should not distract from the real ethical issues.

From an evolutionary perspective, we humans have good reason to be wary of things that seem to be 'unnatural'. Anything out of the ordinary can be dangerous. But the evolutionary origin of that response also guarantees that it will be guided more by emotion than by reason. Witness the reaction last week when Thomas Beatie, from Bend, Oregon, announced his pregnancy on the popular television talk show, *Oprah*.

Beatie, who was born female (and participated in beauty pageants), underwent hormone treatment and some gender-reassignment surgery ten years ago, but retained his reproductive organs. He stopped taking hormones so that he and his wife, who cannot bear children, could pursue artificial insemination. Several doctors turned them down, but last week, the world watched as a baby-faced man with a thin beard and a growing paunch went for an ultrasound: the fetus was a girl. Oprah Winfrey was supportive as she nursed the nervous Beatie through a discussion of his personal realizations. So was the lesbian, gay, bisexual and transgender community. But other reactions were vitriolic, as when MSNBC's Joe Scarborough repeatedly commented that he was "going to be sick". Other such visceral responses were common on message boards and blogs on the Internet, where the situation was often held to be disgusting and unnatural.

And yet, when we consider this story with the reasoning parts of our brains, exactly what was so 'unnatural'? The longing to have a baby? That is a profoundly human desire, whether the prospective parents are male, female or transgendered. Or is it that Beatie has acted on his certainty that he is a man who happened to be born with-

out a Y chromosome? Biologists have found that gender-straddling and gender-switching behaviours are not at all uncommon in the 'natural' world, either for humans or non-human animals (see page 678). True, modern biotechnology has considerably raised the stakes, and is allowing humans to manipulate their biological make-up to an ever-increasing degree. But it hasn't fundamentally changed the game. And its applications, however unsettling they may be to some people, are not, by definition, 'unnatural'.

This same question of 'natural' versus 'unnatural' also emerges this week in a very different context: an online poll that *Nature* started in January on the use of neuroenhancing drugs (see page 674). Respondents were asked to report on their non-medical use of drugs such as modafinil and methylphenidate to improve their concentration. These drugs can have mild effects, not all that different from caffeine (a natural substance) or other stimulants. But somehow the 'unnaturalness' of these drugs makes some people uneasy in a way that caffeine does not. The claim, repeated in many responses to our survey is that using such drugs, or any performance-enhancing drug, makes accomplishments somehow less worthy because they aren't natural. But again, what is 'natural'? Devices such as glasses, hearing aids, pacemakers and artificial hips are unnatural. Yet they are widely accepted as legitimate ways to enhance the human experience. By the same token, if drugs enhance performance on a standardized test, what is so 'natural' about prep courses designed to improve scores?

Ultimately, our visceral concept of what is 'natural' depends on what we are used to, and will continue to evolve as technology does. But in the meantime, we should not allow it to distract us from the rational consideration of deeper and more important ethical issues. In the case of Beatie and his wife, the elemental questions are the health, safety and emotional security of the child. Trying to decide such issues simply by fixating on a fluid and arbitrary definition is, by nature, silly. ■

Ready or not

Transparency and honesty are essential if the genetic-testing industry is to live up to its potential.

Navigenics, a California start-up company with solid backing, launched its flagship product this week. The Health Compass, a US$2,500 genetic test, is being offered to consumers directly, over the Internet. It will scan DNA from a customer's saliva sample for a host of tiny variations and pronounce on the person's risk of developing 18 common medical conditions, including heart attack, prostate cancer and type 2 diabetes.

Why now? Because, as the company says in its corporate literature, "the science is ready. Genetic testing is ready to enter into common health-care practice."

Certainly genetic testing is here, ready or not. In the past year, genome-wide association studies have begun to pour out of labs, linking the blips in our genetic make-up to risks of developing particular medical conditions. Whether people would or could change their behaviour to ameliorate these risks remains unclear. But the ink on the research papers is barely dry before companies unveil commercial versions of the tests. Navigenics is simply the most recent; others include high-profile players such as 23andMe in Mountain View California (see *Nature* **450,** 11; 2007) and the Icelandic genomics company deCODE Genetics. They also encompass smaller operations such as Philadelphia-based Smart Genetics, which last month offered a genetic assessment of the risk of Alzheimer's disease. Rarely have basic discoveries morphed into a commercial product quite so swiftly.

Following almost as quickly are the concerns being raised about the use of such tests. Many of these worries have been around for some time, but now that the tests are here, these discussions have taken on a new urgency. If consumers are to reap the benefits that genetic testing can offer, they need understandable information about the basis, validity and limitations of the tests. One proposed structure for providing this information is a publicly accessible registry into which test-makers would be required to upload data about their tests and the studies that back them. This information should be updated as genetic risks are changed or refined, as inevitably they will be.

Such a registry should be international, harmonizing information in what will doubtless be an industry without borders. This approach seems preferable to stepped-up regulation by agencies such as the Food and Drug Administration (FDA), which — in addition to travelling at the snail's pace of bureaucracy rather than the lightning speed of burgeoning markets — could easily have the effect of driving less-than-desirable players underground, where sub-standard tests will remain as easy to buy as black-market DVDs.

Many critics would say that a hands-off approach by the FDA is irresponsible. But on what basis should genetic tests be treated any differently from others done in government-certified clinical labs? Many of these tests — including almost all of those done by labs in-house, rather than sent out to patients and doctors as kits — are not required to demonstrate clinical utility on a test-by-test basis. For that matter, why should genetic tests be treated differently from medical devices such as MRI scanners, which were left by the FDA to prove their utility and clinical validity to physicians and providers in the marketplace, rather than in pre-market assessments?

It would be naive to suggest that transparency will solve all problems, or to assume that the marketplace will separate the wheat from the chaff with unfailing accuracy and efficiency. But to advocate relatively light regulation does not mean turning a blind eye to the risks of such a strategy. It means taking seriously the presumption that people should be free to inform themselves and make their own choices, and that by doing so they may benefit not just themselves but also the overall pace of innovation. Should it become clear that the system is allowing harm, then enhanced regulation will be appropriate.

It is also worth noting that the scientists who have driven this revolution need to assume a prominent role in ensuring that its benefits are not mishandled. Those who start companies, or advise them, can and must lead the way in ensuring that their enterprises are transparent and valid. In the meantime, online shoppers who buy genetic tests would do well to keep asking themselves whether the science is, indeed, ready. ■

RESEARCH HIGHLIGHTS

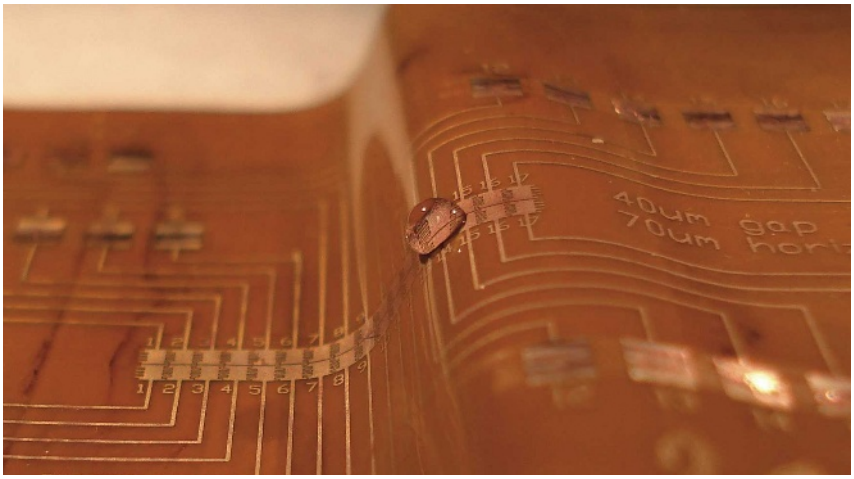
Climbing the walls

Lab Chip doi:10.1039/b801516c (2008)

The ability to steer droplets of various fluids across a suitably wired-up surface with electric fields has a range of potential uses in chemistry and biology. That range may be expanded by the extension of the technique to inclined and vertical surfaces.

Aaron Wheeler and his colleagues at the University of Toronto in Canada bent flexible circuit boards into various shapes and tested their ability to carry drops of different sizes. They were able to propel droplets of up to 8 microlitres up vertical walls and others of up to 50 microlitres along steady slopes.

This three-dimensional extension of digital microfluidics could prove useful for applications in which a droplet has to be bathed in a well of some immiscible fluid and then moved on elsewhere.



M. ABDELGAWAD

MICROBIOLOGY

Fat lazy bacilli

PLoS Med. **5,** e75 (2008)

It is commonly thought that the bacilli found in the saliva, phlegm and mucus of people infected with tuberculosis are active and growing. Michael Barer at the University of Leicester Medical School, UK, Philip Butcher at St George's University of London, and their colleagues report that a significant proportion of the bacilli in this sputum — ones with distinctive lipid bodies in their cytoplasm — do not conform to this view.

These "fat and lazy" bacteria do not replicate, and show greater tolerance to antibiotics than their more active kin. This might reflect an adaptation to the rigours of transmission from person to person, and may help to explain why antibiotic treatment for tuberculosis normally has to be prolonged.

NANOTECHNOLOGY

Tube conductors

Nano Lett. doi:10.1021/nl080302f (2008)

Single-walled carbon nanotubes can be metallic or semiconducting, but are synthesized as an unhelpfully entwined mixture of the two. By using an ultrafast centrifuge that sorts the tubes by density, Alexander Green and Mark Hersam from Northwestern University in Evanston, Illinois, produced a sample of almost exclusively metallic tubes.

Green and Hersam made films from the metallic nanotubes that were transparent and up to ten times more conductive than a film made from mixtures of tubes. The effect of the sorting process also meant that tubes of differing diameters could be easily picked out. Tubes with different diameters were used to produce films with different colours and conductivities. There are applications for transparent conducting materials in flexible circuitry and displays, and even in artificial muscles.

GEOPHYSICS

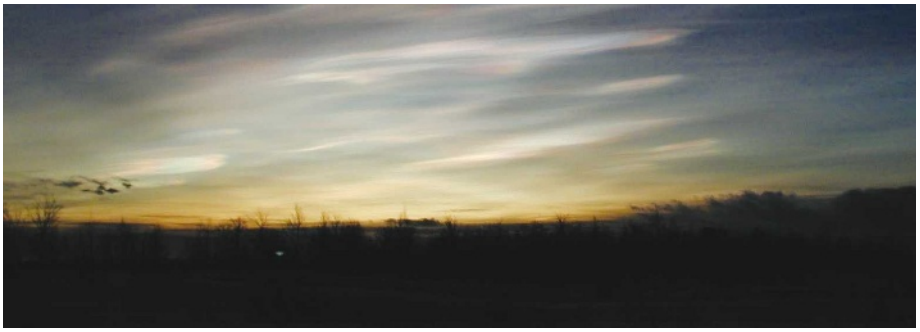
Going deep

Geology **36,** 275–278 (2008)

Geophysicists have captured images of two subduction zones — places where one tectonic plate overrides another — that allow comparisons to be made between them.

The work by Stéphane Rondenay of the Massachusetts Institute of Technology in Cambridge and his colleagues shows subducted slabs behaving differently in the two areas. In central Alaska, where the crust is 15–20 kilometres thick, the researchers observed the sinking slab penetrating to 120 kilometres below the surface. In Cascadia, along the northwestern coast of the United States, the 8-kilometre-thick crust was only seen to a depth of 40 kilometres.

The subducting slabs become invisible to the seismic imaging technique at depth because increasing pressure transforms the rocks into eclogite, an assemblage of minerals that is almost indistinguishable from deeper mantle rocks.



L. POOLE, NASA LANGLEY RESEARCH CENTER

CHEMICAL PHYSICS

Icy lines

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0710129105 (2008)

Cold ice has stripes, according to Erio Tosatti at the International School for Advanced Studies in Trieste, Italy, and his colleagues.

Slicing through an ice crystal exposes dangling oxygen and hydrogen atoms. The crystal structure of bulk ice suggests that these atoms are randomly arranged, but the researchers' computer simulations show that at temperatures below 180 kelvin the cut surfaces will reorganize themselves into alternating rows of hydrogens and oxygens, a state first postulated in 1970.

The findings fit experimental studies of ice. The stripes could affect the surface chemistry of the very cold ice particles that make up polar stratospheric clouds (pictured left).

CONDENSED-MATTER PHYSICS

Spinning together

Phys. Rev. Lett. **100,** 120403 (2008)

Whirling vortices frequently appear in very cold systems of atoms, but usually keep themselves to themselves.

However, vortices in a Bose–Einstein condensate (BEC), an ultracold atomic gas,

may be more sociable, sometimes coming together into 'vortex molecules'. Q-Han Park of Korea University in Seoul and his collaborators modelled what would happen in a BEC composed of both atomic and molecular rubidium. They show that the interactions of the different rubidium species cause vortices in the BEC to form into triplets that look similar to carbon dioxide molecules.

The researchers are now building an experiment to test their calculations. They hope the system will improve the understanding of vortices in other materials, including superconductors.

ARCHAEOLOGY

Excremental advance

Science doi:10.1126/science.1154116 (2008)

The Clovis culture of about 13,200 years ago is the first unmistakeable sign of human presence in the Americas, but there have been tantalizing hints of an older presence. Now, Eske Willerslev of the University of Copenhagen and a large team of his colleagues have found human faeces in an Oregon cave that they carbon date to about 14,300 years ago.

Human mitochondrial DNA from some of the fossil faeces, which were removed from the lowest layer of the Paisley Caves, carries a signature associated with two founding Native American groups. These mysterious pre-Clovis people support findings of non-Clovis cultural artefacts in Chile from about 14,500 years ago.

CANCER BIOLOGY

Long-distance instructions

Cell **133,** 66–77 (2008)

The microenvironment of a breast tumour cell can influence it in a way that leaves it particularly well suited to spread to the lungs.

This may be a general phenomenon, according to Joan Massagué of the Memorial Sloan-Kettering Cancer Center in New York, explaining the mysterious tendency of cancer cells to metastasize to a 'preferred' organ.

He and his colleagues found that in one major type of human breast tumour, cells would invade lungs much more readily than bones if they had responded to the signalling molecule TGFβ in the tumour microenvironment before escaping into the bloodstream.

TGFβ, they found, switches on production of a second signalling molecule in the tumour cell itself. This molecule, ANGPTL4, disrupts the tight connections between cells in the tiny blood vessels that infiltrate lungs — allowing the tumour cells to enter and settle there.

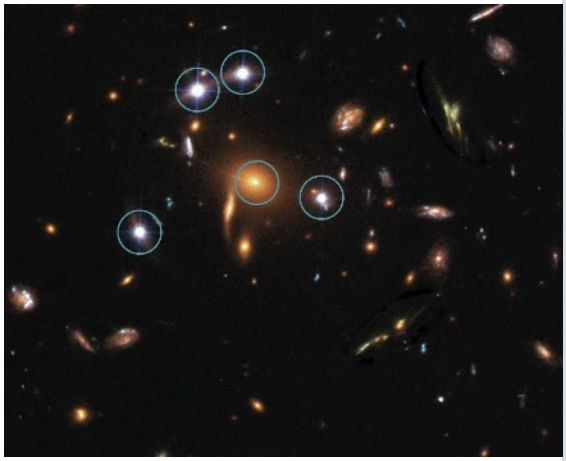
ASTROPHYSICS

Quasar, delayed

Astrophys. J. **676,** 761–766 (2008)

Astronomers have measured the longest time delay yet seen due to gravitational lensing. Earth receives light from quasar SDSS J1004+4112 along five different paths (pictured below), thanks to the space-warping effects of intervening galaxies. Janine Fohlmeister of the University of Heidelberg, Germany, and her colleagues have now shown that one of the three main paths is two years and three months shorter than another, and that the quickest of the paths is at least 6 years shorter than the longest.

This large delay raises the possibility of studying events first seen in one image in more detail when they turn up years later in the others. This could allow the most precise estimates yet to be made of the size of a quasar's central engine.



EUROPEAN SPACE AGENCY/NASA/K. SHARON/E. OFEK

PHYSICS

Shields up

Phys. Rev. Lett. **100,** 123002 (2008)

Forget light, objects might also be made invisible to matter. Xiang Zhang and others at the University of California, Berkeley, report that it is theoretically possible to make a cloak that shields objects from incoming atoms — although the technique would work only at temperatures so low that atoms behave like waves.

The cloak would be made of concentric rings of light, an optical lattice configured so as to change the effective mass and electric potential of passing atoms. The changes would deflect incoming atoms around the object before letting them continue along their original path in the same way that 'invisibility cloaks' deal with photons. So far, the authors say, the experimental possibilities look most promising in two dimensions.

JOURNAL CLUB

Norbert Perrimon
Harvard Medical School, Boston, Massachusetts

A signalling scientist marvels at perfect patterns.

The formation of patterns during animal development depends to a great extent on cells, or groups of cells, sending a specific signal that activates a cascade of reactions in the cells that receive and respond to it. Studies of this process in the fruitfly *Drosophila* have provided many insights into the nature of the molecules involved and the mechanisms underlying cell–cell signalling.

The cell surfaces of almost all animals are decorated extensively with large molecules known as heparan sulphate proteoglycans (HSPGs). These modulate most developmental signalling pathways and comprise protein cores modified by the addition of long carbohydrate chains called glycosaminoglycans (GAGs). GAGs are key to mediating interactions between HSPGs and the molecules that they bind.

Recently, Rahul Warrior at the University of California, Irvine, and his colleagues (*Development* **135,** 1039–1047; 2008) explained the puzzling observation that although HSPGs are required for signalling by the protein BMP in certain tissues, they are not required for BMP signalling during very early fly development.

The authors demonstrate that GAG synthesis does not occur in early embryos because the messenger RNAs that encode two enzymes involved in its construction are not translated. Preventing GAG synthesis at this stage allows an 'activity gradient' of BMP to be generated across the embryo that patterns the dorso–ventral axis of the fly. A few hours later, the GAG enzymes are produced, allowing the modified HSPGs to participate in other signalling pathways.

This study illustrates how temporal control of the synthesis of a ubiquitous set of enzymes is used to modulate the activity of signalling pathways in different tissues.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

FDA to vet embryonic stem cells' safety

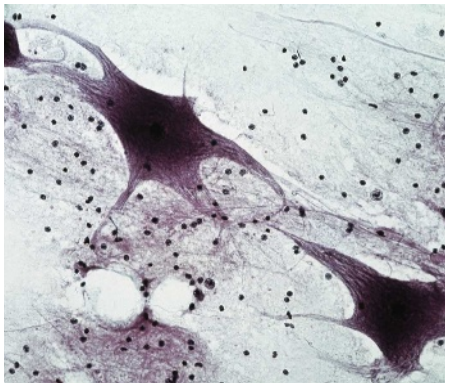
Investors, biotech companies and other stem-cell stakeholders are meeting in Gaithersburg, Maryland, this week for the US Food and Drug Administration's (FDA's) first public hearing on the safety of therapies that use human embryonic stem cells.

The meeting is "a big deal", says Michael Werner, former chief of policy at the Biotechnology Industry Organization in Washington DC. "It could provide clues about what the FDA is thinking in terms of product approvals."

Stem cells derived from human fetal and adult tissues are already being used in clinical trials. But researchers want to use embryonic stem (ES) cells because they show greater capacity to proliferate and differentiate into other cell types. Cells derived from human ES cells have shown dramatic results in treating animal models of disease, but they have not yet been tested in patients, and there are fears that they may carry health risks.

Geron, a biotech company based in Menlo Park, California, plans to start a trial in patients with acute spinal-cord injury this summer using oligodendroglial progenitor cells derived from ES cells. Two other California-based biotech companies, Novocell in San Diego and Advanced Cell Technology in Los Angeles, are also preparing to start human trials using ES-cell-derived products to treat (respectively) diabetes and visual impairment caused by macular degeneration.

The FDA seems to be most concerned about the cells' potential to cause tumours or to differentiate in dangerous ways, and whether the animal safety tests that have been carried out so far provide enough evidence to justify testing ES cells in people, according to FDA briefing documents seen by *Nature*. Another issue concerns how patients should be monitored for signs of problems. Members of the advisory committee have been asked not to speak to the press before the meeting.



M. KAGE/SPL

Geron plans to use human embryonic stem cells to treat patients with acute spinal-cord injury.

The FDA has been looking at these issues for a number of years, according to Michael West, head of BioTime, a biotech in Emeryville, California, and a former executive of Geron and Advanced Cell Technology. "The first time I met with them was in 2001 and they had given it a lot of thought back then," he says.

Human trials will not administer actual ES cells, but cells derived from them, and one of the biggest issues is how to assess the final cell product. One question is how many partially differentiated or undifferentiated cells, if any, are acceptable, and whether undesirable cells can be reliably detected. "You might have cells destined for the spinal cord mixed in with precursor cells destined to make a wisdom tooth," West says. "What happens when you put those cells into the spinal cord?" West says he's not against clinical trials, but he points out that the FDA has a very difficult balance to strike.

No decisions will be made at the meeting, but the process should eventually lead to guidelines from the FDA about required preclinical studies, trial design and patient follow-up, says Marie Csete, chief scientific officer at the California Institute for Regenerative Medicine (CIRM). This will be of great interest to investors, who hope to gain clues about the FDA's expectations. "The greater clarity the FDA can provide around those kinds of rules the easier it is for me to evaluate [ES cells] as an investment opportunity. Right now it isn't clear," says Gregory Bonfiglio of Proteus Ventures, which invests in regenerative medicine companies.

Csete describes the meeting as "a necessary first step" in bringing stem cells to the clinic within a well-regulated environment. CIRM, she says, will expect investigators it funds to implement scientific plans that lead to clinical trials, and she hopes this meeting will indicate what *in vitro* and animal work should be done. "We want to make sure that all the appropriate studies are considered that assure that human ES-cell-derived cell products are optimized for function and safety." ■

Monya Baker

Visit www.nature.com/stemcells for more.

Advanced biofuels face an uncertain future

In trying to assure the market that there is a future for advanced biofuels, the US Congress may have gone too far. Its 2007 mandate for the production of such fuels, intended to convince technologists that a substantial market was guaranteed, may in fact be hampering the technology's development. The targets are so ambitious that many now expect them to be cut back, thus creating an uncertainty just like the one Congress intended to dispel.

Several companies are moving forward with demonstration plants — partially sponsored by the US Energy Department — to produce ethanol from cellulose in corn (maize) stover (the leaves and stalks), wood chips and other plant materials, rather than from the edible part of crops such as corn. But industrial-scale plants remain on the distant horizon, because the technique for efficiently making 'cellulosic ethanol' is still in its infancy. That has many wondering whether the investments are being made to meet the federal mandate, which ramps up to nearly 4 billion litres annually by 2013 and some 60 billion litres annually by 2022.

The biofuels community is watching the Environmental Protection Agency (EPA), which is developing regulations to certify that all of the biofuels in the programme reduce greenhouse-gas emissions by specific levels.

Add the continuing questions about actual production costs for a host of new technologies and long-term questions about the price of oil, and the result is a truck-load of uncertainty, says Wallace Tyner, agricultural economist at Purdue University in West Lafayette, Indiana. "What many people do in this situation is run," Tyner says. The repercussions could be huge if corn ethanol fails as a bridge to cellulosic ethanol. "If oil prices stay high, there's a real temptation to just keep growing corn rather than take the risk on cellulose."

Corn-ethanol production has skyrocketed in recent years, far exceeding the required production. Capacity is expected to hit nearly 40

GETTY

SINGLE-MOLECULE SEQUENCING
Machine is the first of a new generation to hit the market.
www.nature.com/news

Chemical weapons agency shifts focus

Advances in science and changes in the pharmaceutical industry are producing new threats to the international Chemical Weapons Convention, experts warn.

Diplomats from the 183 countries that are signatories to the convention are meeting in The Hague, the Netherlands, on 7–18 April to review the 11-year-old convention, which is generally regarded as a success. To date, implementation of the convention has focused mainly on eliminating existing chemical weapons, although both the United States and Russia will probaby miss their 2012 deadline. The emphasis will soon switch to preventing future production.

One threat to the convention is the pharmaceutical industry, which is building more sophisticated facilities that could easily be converted to produce chemical weapons. And many of these facilities are in countries without established regulatory procedures, according to the Organisation for the Prohibition of Chemical Weapons, the watchdog that oversees implementation of the convention. Such sites should be monitored more closely, it says.

P. PSAILA/SPL

Special facillities have been set up to dispose of chemical weapons.

There are around 4,700 sites the organization calls other chemical production facilities (OCPFs) — which make everything from fertilizers to drugs — even though they do not produce chemicals on the convention's list of controlled agents. "Because of their characteristics and design, a good number of these facilities could quickly convert to the production of toxic chemicals if anyone so wishes. That makes them an object of particular interest," the organization's director-general, Ambassador Rogelio Pfirter, told *Nature*. Only around 10% of these sites have been inspected, and in some countries with large numbers of such facilities, just 1% have been inspected.

In a report released last week, the technical secretariat of the organization says that the agency "remains apprehensive about the present insufficient level of verification" of such facilities. Even though the number of inspections has increased, the effort "still does not provide a sufficient level of assurances for non-proliferation purposes", the report warns. The inspection effort has been hampered by the vast number of OCPFs and the fact that the organization generally relies on host countries to provide information on them, the report says. A new way of selecting sites for inspection is now being implemented, Pfirter says.

A number of countries attending the meeting are expected to agree that the agency should shift its focus from facilities making scheduled agents to OCPFs. However, developing nations such as China, which has a large number of OPCFs, may oppose such a move.

Other key concerns cited by the organization involve the "inexorable march of science and technology" including the growing interface between chemistry and biology and new technologies being developed. Microreactors, for instance, can produce more than 30 tonnes of potential chemical weapons agents a year, including phosgene gas, which was used in the First World War.

"This trend to a more versatile industry means that some of the underlying assumptions we had in the verification system developed in the 1980s may no longer apply," says Ralf Trapp, chemical disarmament consultant and former secretary of the organization's science advisory board.

"If someone were to open a chemical weapons programme today he or she would be looking at those developments, not only at traditional agents," Trapp says. ■

Daniel Cressey

billion litres annually this year and more than 51 billion litres annually in 2009, according to the Renewable Fuels Association, a trade body based in Washington DC. It comes at a time when corn ethanol is facing increased scrutiny. The mandate requires that corn ethanol reduces greenhouse-gas emissions to 20% below gasoline emission levels, but many researchers now believe that it will come up short. This is partly because the resulting higher commodity prices could drive agricultural expansion into wild lands around the globe.

Existing plants were 'grandfathered' in, meaning they won't have to meet the new environmental standards. Given that corn ethanol was capped at 57 billion litres in the mandate, this question affects 10% of the corn-ethanol market. But the entire mandate for cellulosic ethanol — which must reduce emissions to 60% below those of traditional fuels — rides on how the EPA calculates carbon emissions, including potentially indirect effects from land-use changes in the US and abroad. "All eyes are on the EPA to see how they resolve these questions," says biofuels expert Liz Marshall at the World Resources Institute in Washington DC.

"All eyes are on the Environmental Protection Agency at this point."

The industry must also address various infrastructure issues. Corn-ethanol production will soon reach what is known as the 'blending wall' by saturating the current market given the 10% cap on ethanol blends and the scarcity of stations selling an alternative fuel that is 85% ethanol.

David Berry, a principal with the venture-capital firm Flagship Ventures in Cambridge, Massachusetts, says that it will be "incredibly difficult" to meet the mandate for cellulosic fuels. It is likely to be refined and adjusted in the coming years, but this doesn't mean that it will go away entirely, so the technological front runners will be rewarded regardless. "Whether it turns out to be a US$15-billion-dollar market or an $80-billion-dollar market, those are still big numbers," Berry says. He believes that investors will make their move in the next couple of years after the results from various demonstration-scale projects come on line. "We have to figure out who the winners and losers are from a technical standpoint." ■

Jeff Tollefson

SHOWBIZ NEWS

Richard Who?
Evolutionary biologist Richard Dawkins is to follow in the acting footsteps of Derek Jacobi and Kylie Minogue — not to mention his wife, Lalla Ward — by making a guest appearance on the legendary British sci-fi series *Doctor Who*. His late friend Douglas Adams, a *Who* script editor, would have been delighted.

SCORECARD

▲ **Brazilian contraception**
The government aims to safeguard the Amazon by marketing condoms made from sustainably harvested rubber trees in the forest.

▼ **US health information**
A US government-funded health-information search engine called Popline last week blocked all hits relating to the search term 'abortion', only reinstating them after a public outcry.

ON THE RECORD

"It was a big bit of gear and it got pretty warm."

Hafsteinn Jonsson, one of the Swedish team that helped to create the world's first ever 40-gigabit-per-second Internet connection, explains why its recipient, the mother of a colleague, used it to dry her laundry.

M. THOMAS

ZOO NEWS

Strong bond
A loggerhead turtle, badly injured when it washed up on a British beach, has been saved ... by a dentist. The turtle, nicknamed James Bond after his official stranding number (007), damaged his shell on rocks thousands of kilometres from his warmer home waters, and was patched up using a compound more commonly used to protect delicate dental work.

Sources: The Independent, BBC, Wired blogs, The Local, The Times

SIDELINES

Gertrude versus Goliath

An octogenarian physicist is taking on the world's largest electronics companies, which, she claims, have infringed her patent for making blue and ultraviolet light-emitting diodes (LEDs) and short-wavelength laser diodes.

On 20 March, the International Trade Commission, which oversees imports to the United States, announced it would investigate a complaint filed by Gertrude Neumark Rothschild, a professor emeritus at Columbia University in New York City. The complaint names some 30 companies, including Sony, Toshiba and Motorola.

Gertrude Rothschild.

K. TANNENBAUM

Rothschild was awarded a patent for a process used to make wide-band-gap semiconductors in 1993. She had found an affordable way of making the semiconductors' resistivity low enough to make them commercially useful. It eventually allowed green, blue and ultraviolet LEDs and solid-state lasers to become common components in everything from mobile phones to the latest generation of DVD players.

Exactly how companies are producing the diodes is unknown because of trade secrets protection, but there is good reason to believe that Rothschild's process is being used, according to Albert Jacobs, the physicist's lawyer at Dreier, a New York firm. "We believe that her patent covers the only viable commercial process," he says. The companies contacted by *Nature* declined to comment on the case.

Rothschild has taken on the electronics industry before. In 2005, she filed suit in federal court against two other companies: Toyoda Gosei of Japan and and Philips Lumiled. Both were settled out of court. But the trade commission complaint is further-reaching than the previous suits. If the commission finds in her favour then it may order the companies to cease and desist production, or ban the import, of equipment that contains the relevant technologies.

Rothschild says that she filed the complaint in part to raise awareness of women's role in science. "People don't pay too much attention unless there's a financial impact," she says.

This is not the first time that blue LEDs have been the subject of an intellectual-property dispute. In 2004, a Tokyo court awarded Nichia Corporation researcher Shuji Nakamura ¥20 billion (US$180 million) for his role in their development. This was later reduced to ¥840 million after appeal. ■

Geoff Brumfiel

The fraudster returns ...

German academic Hans Werner Gottinger, whose penchant for inventing academic affiliations to promote his career was exposed by *Nature* last year, seems to be up to his old tricks.

This week, the German environmental economist was to run a 15-hour workshop for the Belgian organization Eurosis, at the University of Porto in Portugal. But the workshop was abandoned after the Ostend-based organization, which runs conferences and services promoting modelling and simulation, discovered that Gottinger was not the university professor he claimed to be.

Gottinger told *Nature* last August that he was in 'semi-retirement' (see *Nature* 448, 632–633; 2007). But when he applied to run the Eurosis workshop just a few months later, he claimed to be a professor at the Technical University of Munich in Germany — specifying that he was director of the STRATEC Institute. The university says that it has never employed him and does not host such an institute. It is considering whether to take the matter beyond a formal request to desist.

Eurosis had accepted Gottinger's proposal, which was included in the organization's annual Future Business Technology conference, on 9–11 April. Philippe Geril, secretary-general of Eurosis, says that at the beginning of March, the University of Porto alerted the organization to potential problems. Gottinger withdrew from participation when Geril asked for more details of his affiliation. Gottinger did not respond to *Nature*'s request for clarification.

Gottinger has held many real academic appointments at top universities and research institutes, but frequently used false claims on his CV to acquire them. He has also been accused of extensive plagiarism, including 13 confirmed cases. ■

Alison Abbott

Q&A

Bill Foster

At the age of 19, Bill Foster started a theatrical lighting company that now provides equipment for rock concerts and Super Bowl shows. A PhD at Harvard led him to Fermilab in Batavia, Illinois, where he built electronics for a quark detector. In an election win last month, he entered Congress as the Democratic representative for Illinois. **Eric Hand** caught up with him.

During the election campaign, Republicans ridiculed your science background. Is being a physicist a political asset or liability?
It certainly has been an asset so far. The times that this country gets itself into trouble, particularly of late, are often times when we ignore demonstrable facts. Iraq comes immediately to mind. And I think people understand that Washington would work better if we had less ideology-based discussion and more discussion where the starting points were the facts. And that's the natural stance of the scientist.

Your campaign was more about Iraq and economics than energy and climate science. Do scientists overestimate how important their projects and issues are with the public?
No, I think the public understand the value of science. They just aren't always willing to pay for it right away. The structural problem that science, particularly basic research, has in a democracy, is that it doesn't provide benefits that are visible in the next election. The benefits come 10, 20, 40 years downstream. And so, at least in this country, the politicians tend to systematically neglect basic research the same way they systematically under-invest in education.

Where will your technological knowledge be useful?
Essentially every issue you can name has a technological edge to it, whether you're talking about electronic border fences for immigration, almost anything with armed services — and so on down the list. So there's a benefit to having a scientist many places in Congress, and not just on the pure science-policy side.

Fermilab is a major source of jobs in your district and was protected by your Republican predecessor, retired House Speaker Dennis Hastert. Now the lab is experiencing lay-offs, how can you help it?
The spot to fix the problem will be on the appropriations committee — which you also don't get on as a freshman. I'm relying on the fact that a scientific voice in Congress, and one speaking for the interests of his constituents, is one that is listened to by the leadership.

How do you rate the chances of high-energy physics (HEP) securing supplemental funding this year?
Some mid-year funding relief, through a supplemental appropriation or an attachment to some bill, is one of the avenues we're looking at. My suspicion is it will have happened, or not, by the end of the summer. I've been told by everyone that it will be extremely tough. On the other hand, I like tough challenges.

Your political future may lie in reinvigorating science at Fermilab. What could be done there if the International Linear Collider continues to have technical or political problems?
Fermilab has never had a problem coming up with good ideas for future projects. I was very encouraged to see the flowering of new ideas that is happening there. Most are smaller, intermediate-scale projects that can happen on a faster time scale. I certainly sense a shift in the general opinion of the HEP community that we would do well to no longer put all our eggs in one basket.

You are now one of three PhD physicists in the House, joining Vernon Ehlers (Republican, Michigan) and Rush Holt (Democrat, New Jersey). Are you going to start a new physicist caucus?
I'm stealing Rush's joke — because he congratulated me by telling me how excited he was that we now had a Democratic majority in the physics caucus. However, we've discovered we have to be very careful of sitting together, the three of us, because people start referring to it as the nerds' table. ■

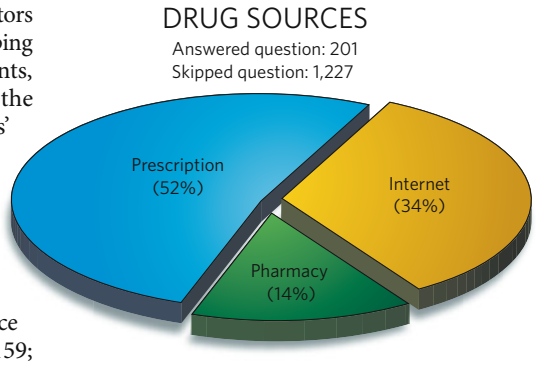
Poll results: look who's doping

In January, *Nature* launched an informal survey into readers' use of cognition-enhancing drugs. **Brendan Maher** has waded through the results and found large-scale use and a mix of attitudes towards the drugs.

The US National Institutes of Health is to crack down on scientists 'brain doping' with performance-enhancing drugs such as Provigil and Ritalin, a press release declared last week. The release, brainchild of evolutionary biologist Jonathan Eisen of the University of California, Davis, turned out to be an April Fools' prank. And the World Anti-Brain Doping Authority website that it linked to was likewise fake. But with a number of co-conspirators spreading rumours about receiving anti-doping affidavits with their first R01 research grants, the ruse no doubt gave pause to a few of the respondents to *Nature*'s survey on readers' use of cognition-enhancing drugs.

The survey was triggered by a Commentary by behavioural neuroscientists Barbara Sahakian and Sharon Morein-Zamir of the University of Cambridge, UK, who had surveyed their colleagues on the use of drugs that purportedly enhance focus and attention (*Nature* **450,** 1157–1159; 2007). In the article, the two scientists asked readers whether they would consider "boosting their brain power" with drugs. Spurred by the tremendous response, *Nature* ran its own informal survey. 1,400 people from 60 countries responded to the online poll.

We asked specifically about three drugs: methylphenidate (Ritalin), a stimulant normally used to treat attention-deficit hyperactivity disorder but well-known on college campuses as a 'study aid'; modafinil (Provigil), prescribed to treat sleep disorders but also used off-label to combat general fatigue or overcome jet lag; and beta blockers, drugs prescribed for cardiac arrhythmia that also have an anti-anxiety effect. Respondents who had not taken these drugs, or who had taken them for a diagnosed medical condition were directed straight to a simple questionnaire about general attitudes. Those who revealed that they had taken these drugs, or others, for non-medical, cognition-enhancing purposes were asked several additional questions about their use. Here's what they had to say:



One in five respondents said they had used drugs for non-medical reasons to stimulate their focus, concentration or memory. Use did not differ greatly across age-groups (see line graph, left), which will surprise some. Nora Volkow, director of the National Institute on Drug Abuse (NIDA) in Bethesda, Maryland, says that household surveys suggest that stimulant use is highest in people aged 18–25 years, and in students.

For those who choose to use, methylphenidate was the most popular: 62% of users reported taking it. 44% reported taking modafinil, and 15% said they had taken beta blockers such as propanolol, revealing an overlap between drugs. 80 respondents specified other drugs that they were taking. The most common of these was adderall, an amphetamine similar to methylphenidate. But there were also reports of centrophenoxine, piractem, dexedrine and various alternative medicines such as ginkgo and omega-3 fatty acids.

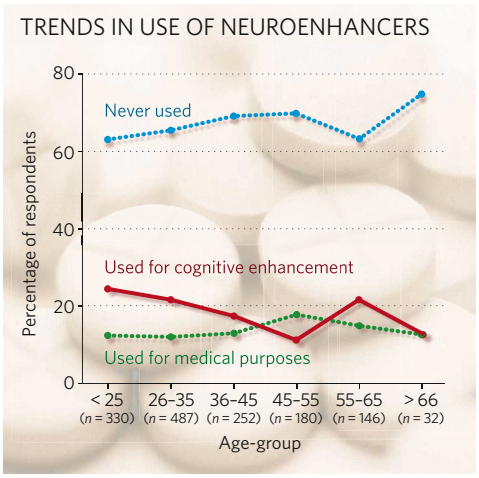
The most popular reason for taking the drugs was to improve concentration. Improving focus for a specific task (admittedly difficult to distinguish from concentration) ranked a close second and counteracting jet lag ranked fourth, behind 'other' which received a few interesting reasons, such as "party", "house cleaning" and "to actually see if there was any validity to the afore-mentioned article".

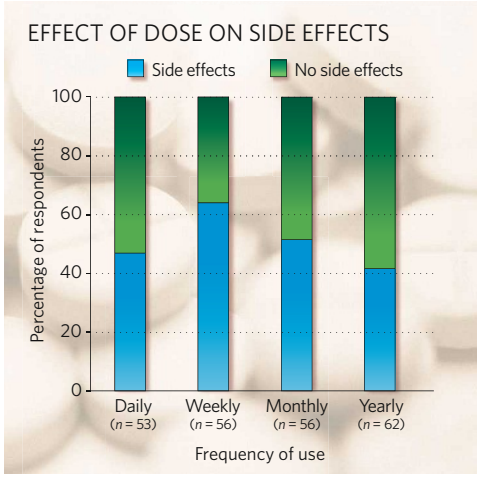
Our question on frequency of use, for those who took drugs for non-medical purposes, revealed an even split between those who took them daily, weekly, monthly, or no more than once a year. Roughly half reported unpleasant side effects, and some discontinued use because of them. Some might expect that negative side effects would correlate positively with a low frequency of use, but that doesn't seem to be the case in our sample (see bar graph, below).

Reported side effects included headaches, jitteriness, anxiety and sleeplessness.

Neuroscientist Anjan Chatterjee of the University of Pennsylvania in Philadelphia predicts a rise in the use of these drugs and other neuroenhancing products and procedures as they become available (A. Chatterjee *Cam. Q. Healthc. Ethics* **16,** 129–137; 2007). Like the rise in cosmetic surgery, use of cognitive enhancers is likely to increase as bioethical and psychological concerns are overcome (see 'Worrying words') and as the products gain cultural acceptance. One difference, Chatterjee says, is that use of cognitive enhancers doesn't rely on training of medical specialists such as surgeons. Internet availability will also greatly accelerate use, he says.

Our poll found that one-third of the drugs being used for non-medical purposes were purchased over the Internet (see pie chart). The rest were obtained from pharmacies or on prescription. It is unclear whether the prescribed





BAD VIBRATIONS
NASA hits a stumbling block.
www.nature.com/news

NASA/MSFC

J. RAE/GETTY

neuroenhancers were diverted from other people's prescriptions, prescribed for different purposes or at different doses for the user. A breakdown of how such drugs were obtained in different countries shows that slightly fewer US users get drugs from the Internet. In the few respondents from Britain that answered this question ($n = 14$), all but one reported the Internet as their source.

All participants who took part in the survey were asked 10 questions designed to gauge their attitudes towards neuroenhancing drugs. Almost all respondents (96%) thought people with neuropsychiatric disorders who have severe memory and concentration problems should be given cognition-enhancing drugs. But perhaps surprisingly, a high four-fifths thought that healthy adults should be able to take the drugs if they want to. And 69% reported that they would risk mild side effects to take such drugs themselves.

When asked whether healthy children under the age of 16 should be restricted from taking these drugs, unsurprisingly, most respondents (86%) said that they should. But one-third of respondents said they would feel pressure to give cognition-enhancing drugs to their children if other children at school were taking them. Morein-Zamir found this coercive factor very interesting. "These numbers strongly suggest that even if policies restricted their use by kids, pressure would be high for parents," she says.

Few studies have looked in depth at the prevalence of this kind of drug use or at people's attitudes towards it. And few data are available on how effective these neuroenhancing agents are or on long-term side effects. The NIDA doesn't fund any such studies, according to Volkow, but the US Department of Defense does. Chatterjee says that he is working with Martha Farah of the University of Pennsylvania on a small study of the effects of these drugs in students.

The most popular of the drugs used by respondents to *Nature*'s poll seem to have fairly mild neuroenhancing effects, says Chatterjee, who calls the massive media interest in these drugs "neurogossip". Nevertheless, the numbers suggest a significant amount of drug-taking among academics. As Eisen's April Fool's prank spread from blog to blog, it was hard to tell who was in on the joke and who was taking the announcement at face value. Although tricking people was a goal, Eisen had been aiming for something so ridiculous that most would chuckle. Instead, he worries that he might have hit a nerve: "I think it did make it less funny because it is actually too real." ■

See Editorial, page 665, and http://tinyurl.com/4huoqr to view and download the survey data and post your own analyses.

Worrying words

Bioethicists have identified four concerns about taking neuroenhancing drugs. *Nature*'s survey respondents confirmed them.

Safety
"The mild side effects will add up to be profound in due course and may even require stronger therapy to control the addiction."
26–35 years old from Nigeria

Erosion of character
"I wouldn't use cognitive enhancing drugs because I think it would be dishonest to myself and all the people who look to me as a role model."
25 or younger from Guyana

Distributive justice
"Morally puts a disadvantage to people without access."
55–65 years old from the United States

Peer pressure
"As a professional, it is my duty to use my resources to the greatest benefit of humanity. If 'enhancers' can contribute to this humane service, it is my duty to do so."
66 or older from the United States

Radio sweat gland — 90 GHz

Sweat ducts in human skin act like an array of tiny antennas that pick up radiation at specific frequencies, according to researchers. The finding might one day be used in medical and security technologies to assess a person's mental state from a distance.

A team of researchers in Israel has shown that sweat ducts pick up radiation at frequencies of about 100 gigahertz — the so-called extremely high frequency or EHF range, lying between microwaves and terahertz radiation. The antenna behaviour is all down to the ducts' curious shape: they thread through the epidermis as regular helices. Filled with electrically conductive sweat, these channels act rather like coils of wire that absorb radiation across the millimetre and sub-millimetre wavelength band.

Yuri Feldman of the Hebrew University of Jerusalem and his colleagues directed a beam of EHF radiation onto the skin of the palms of subjects who had been jogging for 20 minutes, and measured the radiation that was reflected back. They found a strong band of absorption that was not seen before exercise. This absorption gradually disappeared as the subjects rested after jogging (Y. Feldman *et al. Phys. Rev. Lett.* **100,** 128102; 2008). The researchers also found that the reflection signals were proportional to blood pressure and pulse rate — known indicators of physiological stress leading to sweating.

And when the researchers suppressed palm sweating with a synthetic compound that mimics the paralysis of snake venom, inactivating the sweat glands, they found that EHF absorption during exercise was markedly reduced.

The helical antenna array makes skin a kind of biological metamaterial, Feldman's group says, in which the material's response to electromagnetic radiation is determined by structure rather than composition. Metamaterials made from arrays of tiny electrical circuits are being explored for applications ranging from super-lenses to invisibility shields. "Nature has done what is being attempted extensively today in nanophotonics," Feldman says. "This effect might be used for biomedical and homeland-security applications."

Sweating hands have been used in lie detection, but using physiological parameters in 'polygraph' lie detectors is controversial and was strongly criticized in a 2002 report by the US National Academy of Sciences. "Perspiration is related to increases in emotional arousal," says Paul Ekman, a psychologist in Oakland, California, and an author of the academy's report. But he adds that "it can be the consequence of many different mental processes" — not only lying.

So far, Feldman and his colleagues are cautious about whether the idea will work at all, let alone how it might be applied. For example, they need to find the distance at which a meaningful signal can be detected and how long it takes for the signal to register changes in the biometric parameters. "We are just starting our journey in these uncharted waters," says Feldman. ■

Philip Ball

UK satellite firm acquired by European space giant

Europe's largest space company, EADS Astrium, will acquire British microsatellite manufacturer Surrey Satellite Technology Limited (SSTL).

In a deal announced on 7 April, EADS will pay the University of Surrey between £40 million and £50 million (between US$80 million and $100 million) for a majority stake in the company. Currently, the university holds an 80% share.

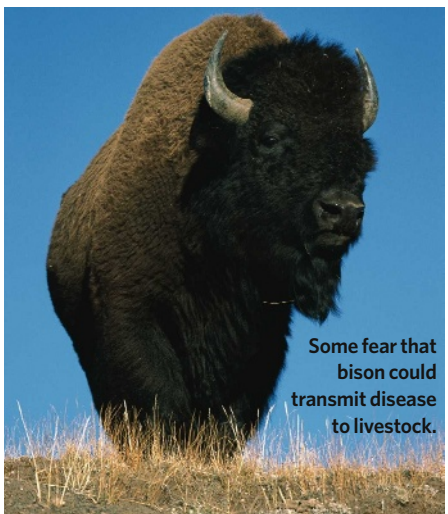
SSTL specializes in small satellites weighing between 10 and 100 kilograms. Its previous clients include nations such as Turkey and China, as well as the European Union, for whom it supplied the first satellite of Europe's Galileo Global Positioning System.

But the company has struggled to expand into larger markets such as the United States (see *Nature* 444, 804–805; 2006). The buyout will help growth, according to Martin Sweeting, SSTL's executive chairman. "This acquisition strengthens SSTL enormously while preserving our unique approach to space," he says.

Brucellosis fears hamper Yellowstone bison plans

A record number of bison foraging outside Yellowstone National Park have been captured and killed this winter as part of a US plan to try to keep the animals from spreading the disease brucellosis to livestock. The bacterium *Brucella abortus* can cause spontaneous abortion in cattle.

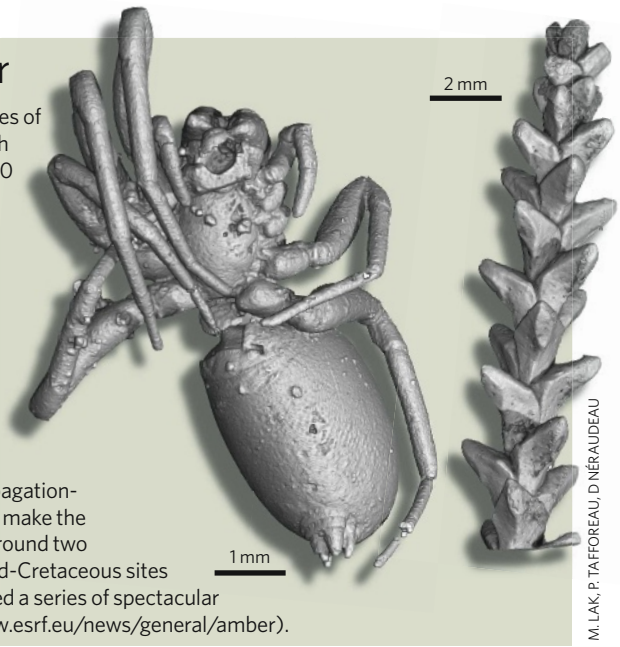
Although the effort has successfully prevented bison–cattle contact, last week the Government Accountability Office reported that the plan has fallen behind in its other goal of allowing more bison to range outside the park. Since 2000, federal and state agencies have struggled to secure additional land for grazing and find a way to deliver the RB51 brucellosis vaccine to bison calves and yearlings.



I. ARNDT/MINDEN PICTURES/FLPA

Some fear that bison could transmit disease to livestock.

An insight into amber



M. LAK, P. TAFFOREAU, D NERAUDEAU

Suspended deep inside opaque pieces of amber, this spider and conifer branch have remained hidden for around 100 million years.

Now they, and some 350 other organisms including wasps, flies and ants, can be viewed thanks to a technique that can see through the fossilized tree resin that holds them.

French palaeontologists at the University of Rennes, in collaboration with the European Synchrotron Radiation Facility in Grenoble, used a synchrotron X-ray imaging technique called propagation-phase contrast microradiography to make the organisms visible. They examined around two kilograms of opaque amber from mid-Cretaceous sites in southwestern France and produced a series of spectacular three-dimensional images (see www.esrf.eu/news/general/amber).

An aerial survey last month revealed that the Yellowstone bison population fell from 4,700 last summer to roughly 3,000.

Tanzania takes steps to save ancient human prints

The world's oldest human footprints are to have a museum built around them in a bid to protect them. The 3.7-million-year-old tracks in an ash bed in Laetoli, Tanzania, are threatened by erosion.

A 15-member government committee approved the 5-year plan, which calls for US$25 million to be raised for the facility.

Archaeologist Paul Msemwa, director of the National Museum in Dar es Salaam who chairs the committee, says that the panel will decide in the next month on the specific remedial measures to protect the footprints. A protective layer placed over the prints is deteriorating (see *Nature* 451, 118; 2008).

Environmental assessments and work are already under way to carve a road into Ngorongoro National Park past the Laetoli site, to provide electricity and a temporary visitor centre, officials say.

Pfizer fails to gain access to peer-review files

A federal magistrate in Massachusetts last week ruled that *The New England Journal of Medicine* (*NEJM*) does not have to comply with a subpoena issued by Pfizer forcing the journal to provide confidential peer-review documents related to the painkillers Celebrex (celecoxib) and Bextra (valdecoxib).

The drug firm had tried to compel the *NEJM* to hand over peer reviews and internal editorial discussions for 11 papers on the painkillers. It argued that these would help it defend the arthritis drugs in lawsuits alleging that they caused heart attacks and strokes (see *Nature* 452, 6–7; 2008).

In his 12-page opinion, Leo Sorokin wrote that the material Pfizer sought seemed relevant on first examination, but that "*NEJM*'s interest in maintaining the confidentiality of the peer-review process is a very significant one … and tip[s] the scales in favor of the *NEJM*."

The judgement comes three weeks after an Illinois judge ruled against Pfizer after it issued almost identical subpoenas to *The Journal of the American Medical Association* and *Archives of Internal Medicine*.

Environment agency reopens doors to libraries

The US Environmental Protection Agency (EPA) last month announced that by 30 September it will have reopened five agency libraries that it recently closed in Chicago, Dallas, Kansas City (Missouri) and Washington DC.

Critics had seen the closures as a move to control information at the agency, and the resulting outcry triggered Congress to allocate $1 million for fiscal year 2008 to reopen the libraries.

The EPA has described the closures as an efficiency move, and said all along that it intends to digitize and make available online the material that was formerly available in hard copy at its libraries. A committee of the Library of Congress is consulting with the EPA on its longer-term library plans.

Correction

Our News story 'Drug markers questioned' (*Nature* **452,** 510–511; 2008) incorrectly stated that Gleevec (imatinib) has been approved to treat breast cancer. Gleevec is used to treat chronic myeloid leukaemia.

SCANDAL!

Some creatures have what it takes to survive long dry spells. How they do this may be revealed in their genes, reports **Erika Check Hayden**.

Sex-starved and still surviving

In the sleepy, tourist town of Woods Hole, Massachusetts, David Mark Welch is looking for a sex scandal, and he knows just where to find it.

He scrapes a small chunk of bark from a locust tree, drops it in a vial and returns to his laboratory in the imposing, red-brick Lillie building at the Marine Biology Laboratory, or MBL, as it is usually known. Once there, Mark Welch places the bark into a dish of spring water. Within hours, it is swarming with bdelloid rotifers, microscopic animals found all over the world and famous for one scandalous feature: they don't have sex. In fact, the organism may have gone without for tens of millions of years. Mark Welch wants to know how they have managed to last so long and is finding some surprising answers in the structure of their genomes and in how they repair damaged DNA.

Bdelloid rotifers reproduce entirely without males: females package a complete copy of their DNA into eggs that develop, sans fertilization, into the next generation. Asexual reproduction certainly isn't unheard of in the animal world: parasitic bacteria force some insects to reproduce without males and female sharks kept alone in captivity have surprised their keepers by giving birth to baby sharks. But an animal that always reproduces asexually is so anomalous that late biologist John Maynard Smith called the bdelloids' mere existence an "evolutionary scandal"[1]. Only bdelloids, and perhaps one or two other animals, have gone for so long without. In this time, they have thrived, adapting to harsh environments and proving extraordinarily resistant to pathogens, seeming to glean the benefits of sex without any of its drawbacks.

Bdelloid rotifers such as *Rotaria macrura* could be the only ancient asexual animals.

David Mark Welch once lived in fear of finding a hypothesis-busting male bdelloid.

D. FONTANETO & G. MELONE

Doubled up

For all its good points, sex does have complications — and not just of the morning-after sort. In the 1970s, Maynard Smith calculated what he called the "two-fold cost of sex". He reasoned that if an asexual mutant suddenly arose in a sexual species, all of its offspring — being female — would be able to reproduce. The male offspring of its sexual counterpart, however, cannot. The asexual mutants would therefore produce twice as many offspring capable of reproducing than the sexual individuals, and would soon dominate the population. Indeed, this is exactly what happens when scientists introduce asexual mutants into cultures of sexual organisms such as yeast, algae or some rotifers.

Yet outside a Petri dish the vast majority of animals are sexual. Why? Most evolutionary biologists think that asexual populations have short-term advantages, but eventually die off without new infusions of genetic variation to help them adapt to change. That's the idea behind prominent explanations for sexual reproduction, such as the Red Queen hypothesis (sex helps defeat pathogens) and Muller's ratchet (asexual organisms can't rid themselves of harmful defects).

Since the seventeenth century, when bdelloids were first spotted, legions of amateur naturalists have trained their microscopes on them. But no one has ever seen a male. Bdelloids' success flies in the face of the idea that asexuality is an evolutionary dead end. More than 400 species of them live all over the world, in all sorts of places: mosses, lichens, bird baths, puddles, sewerage outflows and streams — not to mention tree bark. But their very uniqueness has prompted doubts: surely there must be better evidence of asexuality than our inability to find males?

It doesn't take two: The bdelloid *Philodina roseola* never gets to couple with another of its species.

D. MARK WELCH

Mark Welch first started looking for such evidence in 1989, when he joined Matthew Meselson's lab at Harvard University. Meselson, a molecular biologist, proposed that bdelloid DNA might provide a definitive sexual history. In sexual organisms, such as humans, opposite sexes each package half of their genetic material into sperm or eggs through a process called meiosis. Their DNA is then mixed together in their offspring. Constant mixing over generations tends to homogenize the gene pool, keeping different individuals' genes relatively similar. But bdelloids don't undergo meiosis or exchange DNA, so each individual is free to accumulate her own mutations and pass them on. An asexual population's overall gene pool, therefore, should contain an increasingly diverse hodgepotch of genes. Testing species heterozygosity — the amount of divergence between different individuals' DNA in the same species — might thus be a test for asexuality.

Mark Welch measured heterozygosity in one gene in four bdelloid rotifers and in seven related, but sexual, monogonont species. An average sexually reproducing species is around 1% heterozygous, and so were the monogononts. But the bdelloids were anywhere from 3.5% to 54% heterozygous — a whopping difference. Mark Welch also found that individual bdelloids carried as many as four very different copies of the same gene — something he did not find in any monogonont rotifers[2]. He and Meselson had provided further evidence that bdelloids were asexual, and their methods were later used to cast doubt on other ancient asexuals (see 'Darwinulids exposed'). The small but enthusiastic worldwide community of rotifer biologists was impressed. "That paper showed they are really asexual — they are not just doing it in the dark when you're not looking," says Bill Birky, an evolutionary geneticist at the University of Arizona in Tucson.

> "Bdelloids are really asexual — they are not just doing it in the dark when you're not looking."
> — Bill Birky

Cryptic behaviour

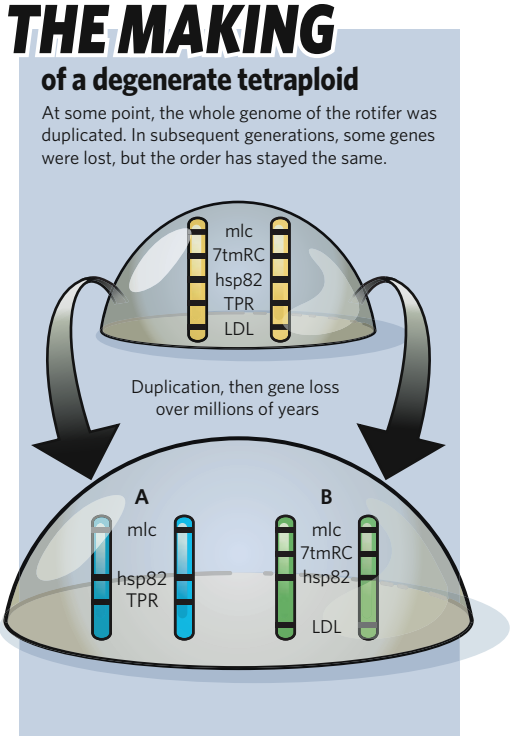
Still, the question remained as to how the bdelloids could maintain such remarkable control. Mark Welch started to wonder whether it was related to another unusual bdelloid ability: when water is scarce, bdelloids enter into a state of suspended animation. The slightest moisture can resurrect them from this state — called anhydrobiosis — bringing them back to life after a drought.

In 2004, Mark Welch set up his own lab at the MBL. Here, he has at his disposal the power of the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, which supports more small-scale projects than other centres. Indeed, the Bay Paul centre could have burned through the sequencing required for Mark Welch's original project in a few weeks, cutting years off his decade-long toil. Mark Welch is now using that sequencing power to look more closely at rotifer genomes. He sequenced DNA in and near the four genes he and Meselson had studied previously. This sequencing expedition painted an interesting picture.

Mark Welch found that a bdelloid's genome is organized in what is called a degenerate tetraploid fashion. Organisms with this genome structure store extra copies of many genes on four separate chromosomes, each of which was originally a member of a chromosome pair that was duplicated some time in the past (see graphic)[3]. Degenerate tetraploids abound in nature — and in the lab; the standard model organism, the yeast *Saccharomyces cerevisiae*, for example, is one.

This tetraploid structure might also explain the mechanism behind another bdelloid superpower, one that has been scrutinized by Eugene Gladyshev. A few years ago, Gladyshev, a graduate student at Harvard University working with Meselson, began studying how the bdelloids could survive and revive themselves from drought. Gladyshev decided to bombard the rotifers with radiation, which affects DNA in ways similar to dehydration. He pummelled specimens of *Philodina roseola*, *Adineta vaga*, and their sexual monogonont relative, *Euchlanis dilatata*, with γ-rays from radioactive caesium. Then he measured how the radioactivity affected each species' eggs.

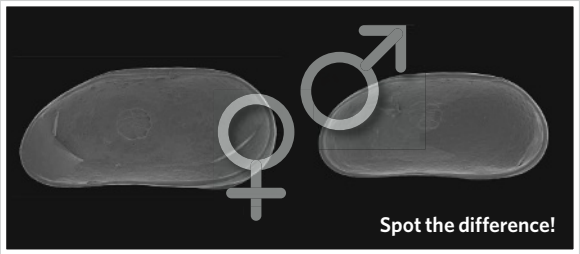
At 200 Grays, the sexual rotifers lost the ability to reproduce — the DNA in their eggs was so scrambled by γ-rays that most of their eggs failed to hatch. But the asexual rotifers shrugged off the dose, hatching just as many daughters as those that weren't irradiated. Gladyshev doubled, tripled and quadrupled the radiation dose — and still, the asexuals' eggs kept hatching. It wasn't until Gladyshev reached a threshold of more than 1,000 Grays that the asexuals finally failed to hatch their eggs. That dose was not only five times higher than the dose endured by the sexual rotifers; it was also a much greater dose than any tolerated by the other animals for which Gladyshev could find similar data[4].



Darwinulids EXPOSED!!

One day in 2003, a biologist collecting specimens stumbled across the nightmare of every scientist who studies ancient asexual organisms: a male[6].

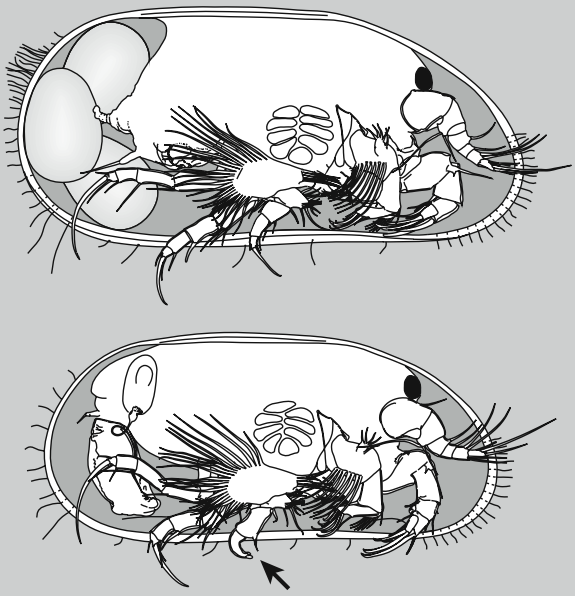
Robin Smith from Lake Biwa Museum in Shiga, Japan, had gone hunting for darwinulid ostracods — tiny freshwater crustaceans long thought to be asexual. He was searching on a forested island called Yakushima, off the coast of southern Japan, with a colleague, Takahiro Kamiya from Kanazawa University. While looking for specimens in a stream near the coast, Smith spotted an ostracod that seemed unusually small. It wasn't until he got back to his lab and dissected the animal that he saw it wasn't just small — it was a male. It had the telltale hooks on its legs, for instance, that male ostracods use to grab their female partners while mating.



Researchers were already dubious about the asexual credentials of darwinulids. They had found low heterozygosity between individuals, suggesting that the tiny creatures could be having sex. But finding a male was a catastrophe for those who believed in the organisms' asexuality. David Mark Welch, who studies asexual rotifers called bdelloids at the Marine Biology Laboratory in Woods Hole, Massachusetts, commiserates: "Throughout my graduate career we lived in fear of reading some report of male bdelloids."

Smith and Kamiya's find raised the question of whether darwinulid males had simply been overlooked in the fossil record because they looked a lot like juvenile females. "Without this fossil evidence, the hypothesis that this group is ancient and asexual doesn't have that much other support," says Smith. He and Kamiya have gone back to Yakushima numerous times, and have only turned up two more males, among thousands more females. So it's not clear whether the males had any role in the population, or whether they were just random curiosities. Nonetheless, the finds raised a caveat about the evidence for asexuality in ostracods — and leaves the bdelloids as perhaps the last thriving asexual animals. **E.C.H.**

REF. 6



Male ostracods have hooks (arrow) to grasp their partners.

Gladyshev and Meselson concluded that the asexual rotifers are "extraordinarily resistant" to ionizing radiation. As the rotifers are unlikely to be forced to endure such extreme radiation in nature, the scientists speculated that the asexual rotifers have actually evolved the ability to patch up their DNA so that they can survive in "ephemeral" environments — places such as tree bark, or puddles, that might be soaking wet one day and bone-dry the next.

To Mark Welch, this idea was very exciting — and not just because it would add to the bdelloids' growing list of superpowers. It also fit nicely with his findings about the tetraploid bdelloid genome structure. When DNA is broken, by drought or by radiation, it can repair itself by using a complementary piece of DNA as a template. And Mark Welch had found two pairs of similar chromosomes, the A and B pairs, in the genome of *P. roseola*. It seemed to him that these pairs could enable the bdelloids' ephemeral lifestyle. Each drought would shatter their DNA. But one chromosome in a pair could mend the damage to its partner by serving as a template.

These super-charged self-healing powers, Mark Welch thinks, could confer certain advantages on the rotifers. Not only might they enable the bdelloids to survive in harsh places where other animals fear to tread, they could also allow the creatures to adapt to change, because their genomes are constantly remodelled during the process of DNA repair. And the self-healing might allow them to rid themselves of invasive elements, such as dreaded "retrotransposons" — genetic invaders that can insert themselves into DNA, copy themselves, and overwhelm a defenceless genome. Researchers have not found any retrotransposons in bdelloids[5]. And the organisms have so far resisted being infected by several of the pathogens that commonly infect *Caenorhabditis elegans* and *Drosophila*. The "rotifers just eat the bacteria", says Mark Welch.

REF. 6

Secret to success

Indeed, the combination of a tetraploid genome and an ability to tolerate drought might be enough to make up for the drawbacks of asexuality, Mark Welch says. And here at last, perhaps, is the solution to why the bdelloids are one of the world's only successful asexual animals: "You have a successful asexual because of this combination of their ecology and their unique physiology," Mark Welch says. "A lot of little things had to be in place, and perhaps nobody else came up with this particular unique set of circumstances."

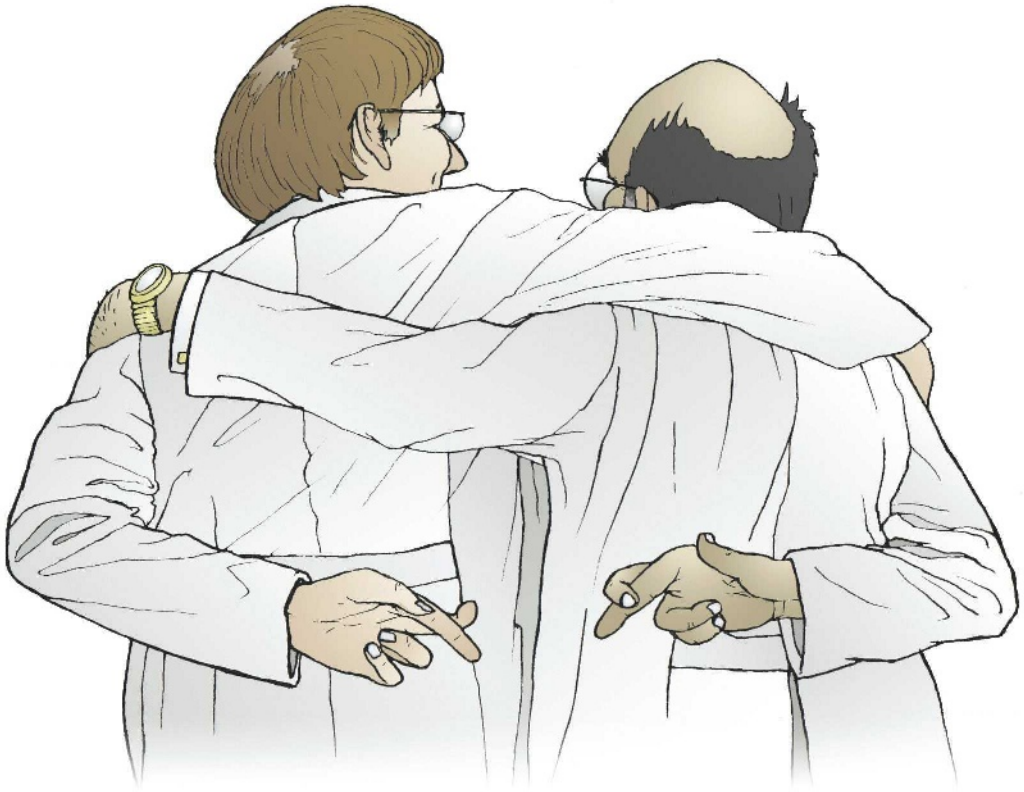
Other biologists are intrigued by the idea too. "The bdelloid work is fascinating, and eventually we may be able to understand why the bdelloids have existed for so long — due to some combination of large population sizes, DNA repair and ancestral polyploidy," says evolutionary biologist Roger Butlin at the University of Sheffield, UK. "I suspect they are all contributing." Yet, he says, "I am not sure why that ends up answering the question of why sex is widespread."

This, of course, is the big question — if the bdelloids are so unique, what are they telling us about the rest of the world? A lot, says Mark Welch. If it's true that bdelloids have survived as asexuals because they can do things like cleanse their genomes of pathogens and maintain genetic diversity, then those may be the very same things accomplished by sex. Perhaps evolution, nature's inveterate tinkerer, has come up with more than one solution for these problems.

Immunity and diversity might not sound like poetic reasons for having sex — you were expecting, perhaps, a paean to the joy of companionship and the thrill of two organisms in union? For the bdelloids, at least, it's hardly missed. ■

Erika Check Hayden writes for *Nature* from San Francisco

1. *Nature* **324,** 300–301 (1986).
2. Mark Welch, D. B. & Meselson, M. *Science* **288,** 1211–1215 (2000).
3. Mark Welch, D. B., Mark Welch, J. L. & Meselson, M. *Proc. Natl Acad. Sci. USA* **105,** 5145–5149 (2008).
4. Gladyshev, E. & Meselson, M. *Proc. Natl Acad. Sci. USA* **105,** 5139–5144 (2008).
5. Arkhipova, I. & Meselson, M. *Proc. Natl Acad. Sci. USA* **97,** 14473–14477 (2000).
6. Smith, R. J., Kamiya, T. & Horne, D. J. *Proc. Biol. Sci.* **273,** 1569–1578 (2006).



J. H. VAN DIERENDONCK

With all good intentions

Collaborations spawn fresh ideas and boost productivity — most of the time. **Heidi Ledford** examines what happens when a working relationship breaks down, and asks how to avoid it.

Break-ups can be painful, as biologist Paul Weldon and chemist Andrew Evans can attest. Weldon spent nine months and filed a lawsuit trying to retrieve a sample he had shared with Evans as part of a collaboration. And Evans estimates that the failed collaboration cost him dearly in wasted time, energy and materials.

Of course, that was by the bitter end of the relationship. At the start, two years ago, things looked a lot more promising. Weldon, a research associate at the Smithsonian Institution's National Zoological Park in Washington DC, had just isolated a compound called bovidic acid from the oily skin secretions of the gaur (*Bos frontalis*), an endangered ox. He suspected that it was the animal's natural mosquito repellant, but he needed a chemist to create a synthetic version of the molecule and confirm that he had found the right compound.

Weldon asked Evans, then at Indiana University in Bloomington to team up with him, and he sent Evans a sample of bovidic acid. The two corresponded by e-mail, and Weldon travelled to Evans's lab. Neither saw the need to formalize their relationship in writing — they had collaborated with researchers before and such measures had never felt necessary. It was a decision they would both come to regret.

The partnership slowly fizzled as the two collaborators' interests diverged. Then conflict arose over who owned the samples. "Collaborations always start off in a woolly way," says Evans, who is now at the University of Liverpool, UK, "and it either takes off and goes great or one of you goes off in another direction — and that's what happened here."

"I feel like I've been hit by a truck," Weldon says. "I'm going to be very circumspect from now on." But Evans has sent material to *Nature* suggesting that another, earlier collaboration of Weldon's, relating to bovidic acid, had also ended with disagreement between the collaborators. Weldon says that of about 30 collaborations he has entered, only those two have ended under difficult circumstances.

Teaming up

More and more researchers are entering into collaborations, often with multiple members and in distant locations. Among the top 200 research universities in the United States, science and engineering publications involving multiple institutions increased 48% between 1988 and 2001 (ref. 1). And between 1990 and 2000, the proportion of publications from international collaborations nearly doubled, accounting for close to 16% of all scientific articles published in Thomson Scientific's ISI Web of Knowledge[2]. The rise of interdisciplinary research and the ease of long-distance communication have encouraged this trend, along with funding agencies that earmark grants for collaborative projects.

Many of these collaborations are productive, generating new friendships, data and ideas. But researchers often underestimate the effort it takes to nurture a successful collaboration, or fail to anticipate the many ways one can go sour. "It is astonishing how little communication there is in the scientific community concerning planning of the project and talking about who is doing what," says Ulrike Beisiegel, an ombudsman for the German Research Foundation in Hamburg. Without such planning, and a written agreement, scientists can find themselves in testy exchanges over paper authorship and data ownership. At best, these are rocky patches that can be smoothed over; at worst, such conflicts can lead to abandoned projects, wasted research funds and lawsuits.

Weldon spent the next nine months fighting to get the bovidic-acid sample back, not because it was valuable but "out of principle", he says. Richard DiMarchi, then the chair of Indiana's chemistry department, urged Weldon and his collaborator to work it out.

Weldon hired a lawyer and, last September, sued Indiana University for possession of all bovidic-acid samples related to the

collaboration. Because the bovidic acid had been purified at Indiana University, the university felt that it could have a claim to ownership, says Beth Cate, associate general counsel of the university. But "in the end we decided it would be best to be done with this matter", she says. "I think there were some misunderstandings here that led to some heightened distrust or emotion." Two months after Weldon filed the lawsuit, the university told him that they would return the sample, and he dropped his claim.

This whole story runs counter to the idea that scientific collaborations are advantageous. In a 2006 analysis of publications, pharmacologist William Figg at the National Cancer Institute in Bethesda, Maryland, showed that the number of citations earned by biomedical articles in elite journals rises with the number of authors on the publication[3]. Another study found that biotechnology papers with at least one international co-author tend to have a higher citation rate than those with a national co-author at a different institution[4]. University administrators and funding bodies sometimes use such studies to promote the virtues of collaboration in the hope of increasing the impact of the research they pay for. Some funding agencies, including the US National Science Foundation (NSF) and the European Commission, have set aside money specifically for interdisciplinary or international collaborations. The European Commission's Seventh Framework Programme, for example, has set aside €32.4 billion ($51 billion) for international collaborative research projects.

L. TODD/DUKE PHOTO

"There are unintended consequences when large funding agencies force people to collaborate."
— Jonathon Cummings

But Jonathon Cummings, professor of management at Duke University in Durham, North Carolina, questions whether this type of forced collaboration is as productive as the figures initially suggest. "There are unintended consequences when large funding agencies force people to collaborate," he says. "If you say we're not going to give you money unless you bring in two to three other universities, it's forcing their hand." Some sociologists have pointed out that it is inaccurate to measure the success of collaborations by counting up publications and then seeing how many stemmed from teams. It overlooks those collaborations that collapsed before papers were written and therefore skews the data towards successful efforts.

In an attempt to reduce this bias, last year Cummings and his colleagues surveyed participants in 491 NSF-funded research collaborations, more than half of which spanned multiple universities. They collected data on all the projects that were funded — including less successful collaborations — and measured what they produced in terms of 'knowledge outcomes', such as patent applications, conference presentations or published articles. The study suggested that projects involving multiple universities produced fewer knowledge outcomes than those involving a single institution[5]. Cummings says that this is because some of the collaborators failed to plan sufficiently for the challenges of coordinating research across the disciplinary and geographical boundaries, so they either wasted time and effort or their collaborations broke down. "When you come together in an attempt to get money but don't really think through how you're going to work together, it ends up backfiring in terms of lower outputs and lower productivity," he says.

Scientists have a rosier view. They often compare their collaborations to romantic relationships and, like true romantics, they feel that these relationships are built on trust. They prize the ability to form spontaneous, unstructured liaisons — ones that spring from chance encounters at a meeting or coffee room — and enjoy the flush of generating new ideas with like-minded colleagues. "Entering collaborations is much like dating," says Neil Smalheiser, a neuroscientist at the University of Illinois, Chicago, who co-founded the *Journal of Biomedical Discovery and Collaboration*. "Some people obsess over it, whereas some proceed blithely and blindly without considering long-term ramifications at all."

The consequences of a failed collaboration cut just as deeply as a failed affair. The most common source of conflict is over publications, which are often the first time that anything is written down. This can raise a host of thorny questions about when data are ready for publication and — the point on which careers are made or broken — who should be on the author list and in what order. The US Office of Research Integrity in Rockville, Maryland, has reported that it receives many misconduct allegations about researchers who have published some aspect of a joint project without crediting a former collaborator, but that these cases are not within its remit. Academic institutions are left to decide whether to launch investigations.

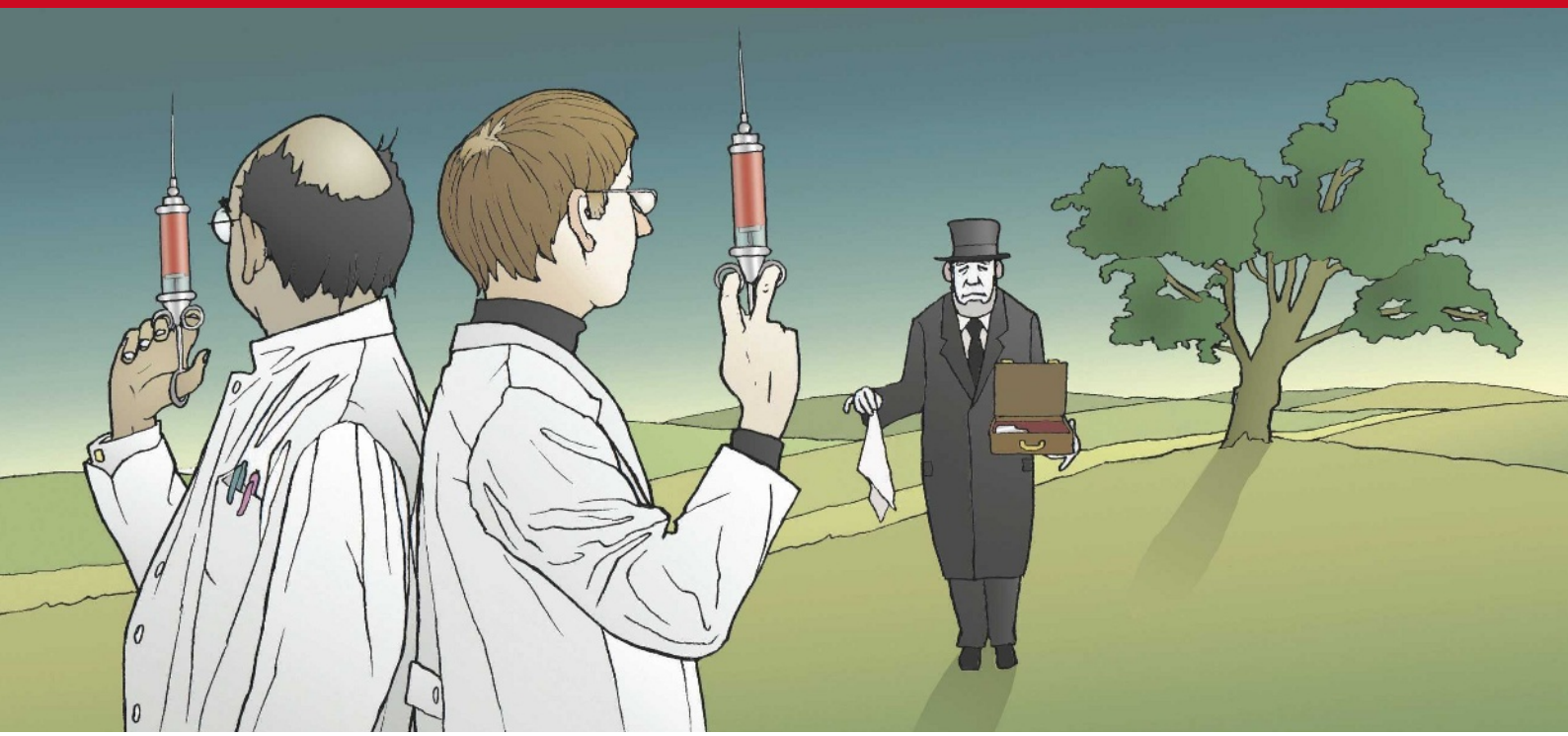
Data ownership is another common source of disagreement, particularly when valuable intellectual property is discovered as part of the work. "People and institutions can get greedy," says Diane Sonnenwald, a sociologist at the University of Gothenburg in Sweden who has studied scientific collaborations. "Research has shown that it's easiest to address these issues before there's real intellectual property to argue over". Beisiegel says that she advises scientists to resort to a lawsuit if a collaborator has threatened their intellectual property. But some researchers shy away from such battles. "It's always a question of how much time and energy you want to spend on these things," says Wesley Shrum, a sociologist at Louisiana State University in Baton Rouge. "Some battles are worth fighting, and some are not."

Write it down

Among those who study collaborations, the solution is clear: potential partners should get everything in writing at the start. Written 'pre-collaboration agreements' can spell out division of labour, data ownership and who will be on the author list. They are sometimes dubbed 'collaboration prenuptials' or 'prenups' in reference to the legal documents that some couples sign before getting married, delineating how property and wealth will be split in the event of a divorce. But many scientists baulk at the idea of a prenup, saying that it can be awkward to press a new collaborator into a written contract before materials and ideas are exchanged, and before knowing where the research will lead. "We recognize that using scientific prenuptials goes against the informal norms of science," wrote Howard Gadlin, ombudsman for the US National Institutes of Health in a 2002 article on responsible research conduct[6]. "But we have seen the damage that can be caused, both scientifically and personally, when scientists at the NIH overlook questions such as these in their enthusiasm to launch an intellectually exciting collaboration."

"I'm going to be very circumspect from now on."
— Paul Weldon

Participants in large, formal collaborations often use their grant applications as a form of prenup, because it forces them to decide precisely who will do what. Many universities have template agreements for researchers to use before establishing a collaboration with industry, but such agreements are not required and often not even encouraged for collaborations



J. H. VAN DIERENDONCK

within academia. "We don't see a lot of those at all," says Peggy Fischer, associate inspector general for misconduct investigations at the NSF. "Maybe we don't see them because they work, and maybe we don't see them because they don't exist."

A rough outline

Smalheiser argues that written prenups are not always necessary, particularly when collaborations are with people one already knows and trusts. "I don't see formal agreements as something practical to pursue," he says. "Especially because the nature of the research and the nature of the collaboration may be continually evolving." He has developed a checklist of issues to hash out early in the collaboration, such as authorship and data sharing[7]. The checklist is intended to raise important issues for discussion up front and, although it can be used as the basis for a prenup, it does not have to yield a formal agreement, Smalheiser notes. Even sketching out a rough outline of the collaboration can be useful, and updating the agreement as conditions of the collaboration change (see 'The collaborators' prenup').

Long-distance relationships can be particularly testing, in science as in romance. As a general rule, the further away a collaborator is, both in discipline and in geography, the more prone the collaboration is to conflict, says John Walsh, a sociologist at the Georgia Institute of Technology in Atlanta. Researchers at different institutions are more likely to face conflicts over schedules, for example, and disciplines may have differing customs for authorship. "There is pressure now both to have larger collaborations and to have remote collaborations that incorporate resources from different places," says Walsh. "You're more likely to be faced with difficulties that you wouldn't face with a local team of modest size." Sorting through the wreckage of a crumbled international collaboration can be complicated by conflicting international policies and laws (see Commentary, page 686). "When you were collaborating with somebody down the road or in the state next door, everybody was operating by the same rules," says Nicholas Steneck, director of the Research Ethics and Integrity Program at the University of Michigan in Ann Arbor. "Now all of a sudden you've got collaborations with labs around the world, and there aren't the same rules and understanding."

The key, says Cummings, is not to break up collaborations but to allow more time and money for management. This might involve hiring staff to help coordinate the programme. In response to his study, the NSF began requesting management plans (which are effectively compulsory prenups) in some of their programmes, particularly interdisciplinary ones. "They've realized that it's one thing to write a proposal for some grand idea," says Cummings. "It's another to spell out who's going to do what and how you're going to do it." The European Commission also requires management plans to accompany its Seventh Framework Programme funding for collaborative projects. But "it's still not a common occurrence across all funding agencies," Cummings says.

Paul Jeffrey of Cranfield University, UK, suggests that funding agencies could also boost collaborative productivity by promoting long-term collaborations, rather than favouring new partnerships. "You can't lock two strangers in the room and expect them to get along, and you can't do that with scientists either," he says. "The overhead of setting up a productive relationship is very high."

Weldon and Evans's overheads were certainly steep. In the battle over bovidic acid they lost time, money and, perhaps most importantly, trust. Weldon now has the bovidic-acid sample back. And although Evans says the sample is intact, Weldon says he is still a little nervous about opening it up and wants to wait until a colleague is present. "I know it may sound a little paranoid," says Weldon. "After all of this, I just want to be careful." So does Evans. Next time, he says, "I'd do everything by the book. You must get the relevant documentation, it's absolutely imperative". ■

The collaborators' prenup

Ten questions to discuss before starting a collaboration.

- What do we expect to get out of this?
- Who is going to do what and by when?
- Who will have access to our data?
- Who will give public presentations, and how much data will they reveal?
- How will we assign authorship?
- How will we decide when to publish?
- Who owns the intellectual property?
- Will we share our reagents with other labs?
- What happens if one of us leaves the project?
- What happens if one of us wants to form a separate, but related, collaboration with another lab?

Adapted from: NIH Office of Ombudsman

Heidi Ledford writes for *Nature* from Cambridge, Massachusetts.

1. www.nsf.gov/statistics/nsf07320/
2. Wagner, C. S. & Leydesdorff, L. *Res. Policy* **34,** 1608–1618 (2005).
3. Figg, W. D. *et al. Pharmacother.* **26,** 759–767 (2006).
4. Frenken, K., Holzl, W. & de Vor F. *J. Eng. Tech. Manag.* **22,** 9–30 (2005).
5. Cummings, J. N. & Kiesler, S. *Res. Policy* **36,** 1620–1634 (2007).
6. Gadlin, H. & Jessar, K. *Preempting Discord: Prenuptial Agreements for Scientists.* The NIH Catalyst (May-June, 2002).
7. Smalheiser, N., Perkins, G. A. & Jones, S. *PLoS Biol.* **3,** e217 (2005).

See Editorial, page 665.

CropLife still committed to assessment's original aims

SIR — Your Editorial 'Deserting the hungry?' (*Nature* **451,** 223–224; 2008) discusses CropLife's decision to withdraw from the World Bank-led International Assessment of Agricultural Science and Technology for Development (IAASTD). As president of CropLife, I wish to confirm that the decision was not taken lightly, given our commitment to agriculture, development and sustainability.

We have been actively involved in the assessment since its inception and we remain committed to its original aims. These were to reduce hunger and poverty, to improve rural livelihoods and to facilitate equitable, environmentally, socially and economically sustainable development through generation, access to and use of agricultural knowledge, science and technology.

Unfortunately, we do not believe that the current draft assessment adequately reflects the role that modern science and technology, and in particular our own industry's technologies, have played in supporting agriculture. In our view, the IAASTD's treatment of biotechnology, crop-protection chemistry, the importance of intellectual property and the role of the private sector has been superficial and negative. Also missing is a vision of science and technology's future contribution to all types of agriculture.

The Editorial focuses on biotechnology, one technology among many that will enable farmers to select the right tool at the right time. We agree that there is no single solution, which means that a range of technologies, modern and traditional, need to be available and that valid, real-world solutions should not be ignored.

We support initiatives to help the rural poor and hungry by providing balanced information about the most appropriate technologies available to improve crop quality and protect the environment. We encourage your readers to look at other reports, such as the *2008 World Development Report* from the World Bank (http://tinyurl.com/6h8bz4) or the Food and Agriculture Organization's *2004 State of Food and Agriculture Report* (http://tinyurl.com/63kr62). Both of these highlight the importance of technology in achieving a productivity revolution, particularly for smallholder farming.

CropLife remains an open and engaged partner with the World Bank and all stakeholders — including governments, international institutions and civil society — to ensure that the original commitments of the IAASTD are honoured.

As stated in the Editorial, we continue to be open-minded about the IAASTD, and we hope that the final report will have corrected the inadequacies of earlier drafts. But it would be counterproductive for us to endorse the current draft.

Howard Minigh
CropLife International, Avenue Louise 143, B-1050 Brussels, Belgium

Italy must invest more in science and technology

SIR — Silvio Berlusconi and Walter Veltroni, the leaders of the two major political parties taking part in Italy's elections on 13 and 14 April, have been debating issues critical for the future of the country's economy — with one crucial omission. Neither of these has mentioned increasing investment in science, technology and education.

Europe and North America realize that their twentieth-century economies, based on manufacturing industry and commerce, are rapidly disappearing under the pressure of emerging economies and globalization. The economy of the twenty-first century will hinge on investment in science and technology — as the US Congress was advised in a 2007 report by the US National Academies, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future.* The European Union has come out with the Lisbon strategy, intending to become "the most dynamic and competitive knowledge-based economy in the world", by investing 3% of gross deomestic product in research by 2010 (http://tinyurl.com/59phnh).

Italy is not following the Lisbon strategy and today is one of the European countries investing least in research (less than 1.5% of GDP — in contrast to 3.86% by Sweden, 2.51% by Germany, 2.13% by France and 1.73% by the United Kingdom, for example (*Nature* **451**, 378; 2008). Italy has only 2.7 scientists per 1,000 of the population, whereas the European average is five and the US average is six. Without more investment, the Italian economy will certainly suffer, to the detriment of the coming generation.

The sparse attention to science in the Italian election campaign reflects a lamentably low interest in science nationwide. Some believe that the Italian university system is not competitive, so no more money should be spent on it until appropriate reforms have been carried out. But reform will not be possible without a sustained increase in research investment. At present, the research budget covers only staff salaries and there is no scope for encouraging the best scientists with increased funding.

Perhaps the scientific community in Italy has failed to communicate effectively the importance of science in the skilled employment and innovative enterprises that are key to our future prosperity. Nonetheless, our national administrators must take this on board and make investment in research and technology the highest priority.

Ivano Bertini*, Silvio Garattini†, Rino Rappuoli‡
*Center for Magnetic Resonance, University of Florence, 50019 Sesto Fiorentino (FI), Italy
†Mario Negri Institute, 20157 Milan, Italy
‡Novartis Vaccines and Diagnostics, via Fiorentina 1, 53100 Siena, Italy

Cleaning up the final phase of the fossil-fuel industry

SIR — Neil Wilson is concerned about technologies that accelerate a net release of fossil carbon to the atmosphere, as he says in his Correspondence (*Nature* **451**, 768; 2008) about our Letter (*Nature* **451**, 176–180; 2008). My co-authors and I share his concern. Fossil fuels will supply society at some level for at least another half-century, given current economic imperatives, although clean-energy cycles that involve zero-carbon emission will ultimately be nuclear in nature, with fusion in the form of solar energy and fission in the form of geothermal energy.

It is therefore crucial to clean up the fossil-fuel industry quickly and economically. Carbon capture and storage are, in theory, deployable. In practice, these are hindered by the enormous number of effective retrofitted storage projects that are needed, and by economic models that allow fossil-energy recovery processes based on existing technologies to be very profitable.

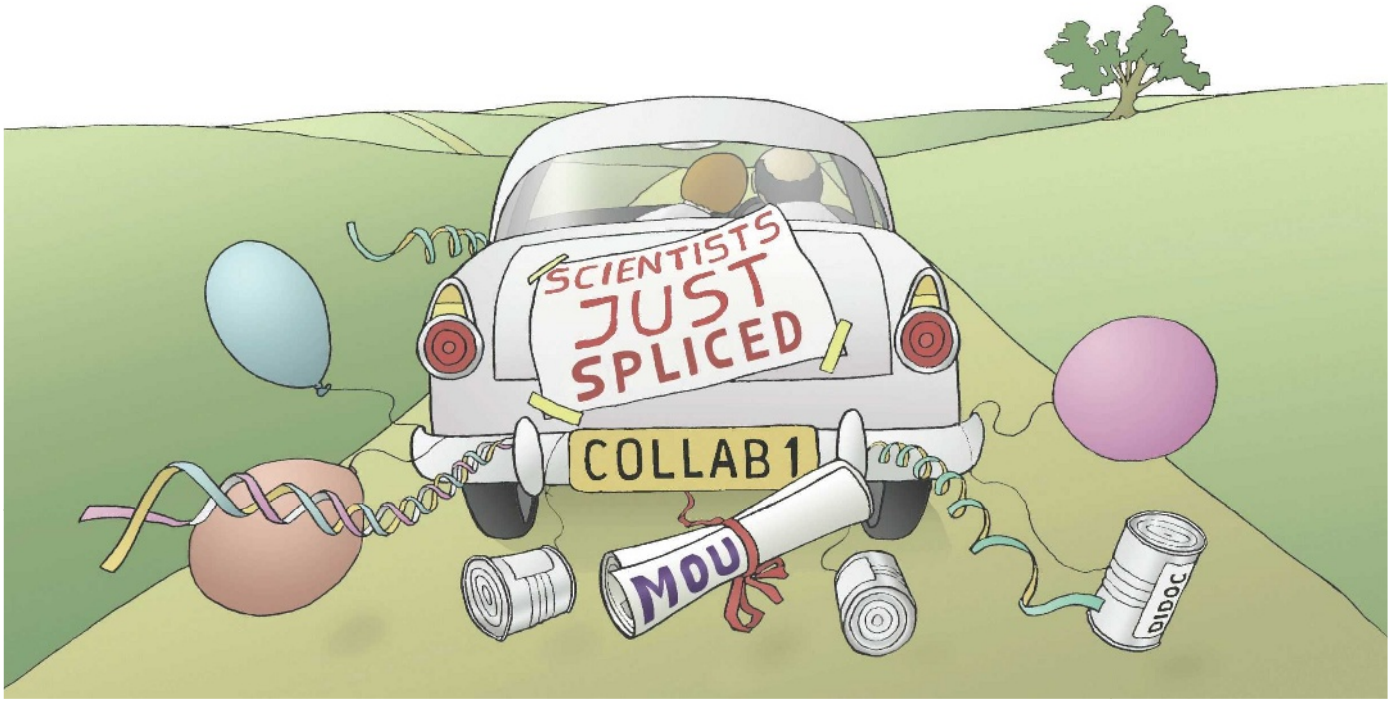
Our Letter is about deep subsurface biodegradation of crude oil and the nature of the deep subsurface biosphere. It raises the possibility that acceleration of this process could provide methane, or even hydrogen, from spent oilfields or as an alternative energy vector for heavy-oil energy recovery. Although this approach remains speculative, it could provide natural gas or hydrogen with low energy and water input, instead of environmentally expensive heavy oil.

Methane and hydrogen are preferable to coal and heavy oil as fuels. The technology — if feasible — could be introduced by the oil industry without major changes in infrastructure or work practices.

This attempt to clean up the final phase of the fossil-fuel industry quickly is one justification for our work at a time when everyone would prefer clean, zero-carbon emission energy to be readily available. Changes in human imperatives call for bold decisions by shareholders, voters and legislators to force rapid and convincing changes in the energy industry.

Steve Larter
Petroleum Reservoir Group, Department of Geology and Geophysics, University of Calgary, Calgary, Alberta T2A 1N4, Canada

COMMENTARY



J. H. VAN DIERENDONCK

Investigating international misconduct

The rise in cross-border collaborations is making it more difficult to police misconduct. **Christine Boesz** and **Nigel Lloyd** argue for a framework to examine allegations and hold researchers accountable.

Research misconduct is an issue the research community cannot afford to overlook. Data fabrication or falsification, plagiarism and other unethical behaviors hit the headlines because they damage the scientists involved and erode the public's faith in the research that its taxes support.

Therefore countries, universities, and other research entities must take seriously allegations that arise and hold researchers accountable. One way to do so is by developing and enforcing formal policies and procedures for handling misconduct allegations. Many countries such as the United States, the United Kingdom, Canada, Denmark, Germany and Australia have such formal policies. Many others do not, including some that are major sponsors of research. As multinational, cross-disciplinary research becomes increasingly prevalent and desirable this patchwork presents problems.

When misconduct allegations are raised within international collaborations, differences within and between national policies create practical challenges that must be resolved fairly and objectively. For example: which country or countries conduct an investigation? Should one country assist another with its investigations? Who do you contact in what country? Worse, what happens when two relevant national policies are at odds?

Research funding agencies such as the US National Science Foundation (NSF) have already faced some of the challenges sketched above. In one case, the NSF Office of Inspector General (OIG) received an allegation of material inaccuracies in a final report of a project intended to facilitate collaboration between a US principal investigator and a foreign scientist. The foreign scientist denied participating in the project and the principal investigator could provide no evidence of the collaboration. After initially assisting in the investigation, the foreign researcher suddenly stopped acknowledging OIG e-mails and refused to provide a sworn statement. The OIG could not compel the non-US researcher to respond to its inquiry and was thus unable to pursue investigative action against the American scientist.

When international students or professors who are accused of research misconduct return to their home countries, resolving allegations becomes more complicated. Investigators often lose touch with such individuals, who are frequently able to continue their studies or find academic positions in their native countries. Similarly, people against whom a finding of research misconduct is made often re-enter academia in a different country where no one knows their history.

A long shadow

Occasionally, international research misconduct receives so much publicity that all scientists involved are forever linked to it. One such instance involved Woo Suk Hwang, the researcher at Seoul National University in South Korea that *Time* magazine named one of its 'People Who Mattered' in 2004. Hwang and his many co-authors published two acclaimed papers — in 2004 and 2005 — claiming to have cloned early-stage human embryos and to have customized stem-cell lines. The papers were later retracted, and Hwang was found guilty of manipulating images, fabricating data, misusing research funds and unethically recruiting donors. He was dismissed from his post, banned from working in a public position for 5 years, and his retirement benefit was halved. Four other professors from the university were suspended and two others had their wages cut.

In one of the related misconduct investigations in the United States, the University of Pittsburgh, Pennsylvania, found Hwang's co-author Gerald Schatten innocent of research misconduct and recommended no disciplinary action. Nonetheless, the university chided Schatten for not assuring the accuracy of a manuscript, and failing to ensure that all co-authors approved the manuscript before submission.

International collaborations that go so awry understandably cast a lasting shadow over interactions between foreign institutions. Mechanisms of investigation — such as those that helped to resolve the Hwang case — are crucial to enabling researchers to initiate and maintain such collaborations.

We are involved in moves to help address these challenges at the Global Science Forum of the Organisation for Economic Co-Operation and Development (OECD).

Building bridges

Last December saw the inaugural meeting of the OECD Co-ordinating Committee for Facilitating International Research Misconduct Investigations*, which we are proud to co-chair. Sixteen participants represented 14 countries and international bodies — Austria, Canada, Denmark, the European Commission, Finland, France, Japan, Korea, the Netherlands, Norway, Portugal, the United Kingdom, the United States and the OECD†. This expert group was created in response to the OECD report‡ recommendation that, "interested countries are encouraged to undertake an international dialog among national practitioners".

The committee's role is advisory, instructive and awareness-raising. We do not have the authority to impose obligatory policies or procedures on any national government or governing body. We can make no binding legal decisions, nor compel any nation to embrace any decisions implied by committee discussions. In reality, we would not wish to impose a given system on our member countries. We much prefer a nation to evaluate its own policy and determine for itself whether that policy serves its intended purposes and how it interacts with those of other national entities.

Attendees discussed practical issues related to international research-misconduct investigations. We talked about harmonizing existing national systems, sharing information, establishing networks, promoting cooperation in dealing with allegations, developing generic models of misconduct-related documents and broadening the committee membership. The first meeting focused on these fundamental topics to open a dialogue between nations, many of which have very different ways of handling investigations.

We agreed that harmonizing the research-misconduct systems of different countries is unlikely given each nation's unique legal and administrative systems. So we agreed to try to come up with a practical set of general principles directly relevant to international research-misconduct investigations, against which countries can evaluate their policies. Some of the principles discussed were: the need for a fair trial; maintenance of confidentiality during the process; protection of informants, accusers, accused and innocent third parties; and agreement between countries to co-operate.

We identified the need for a central database of national research-misconduct policies and for contact details of those responsible for investigations. The absence of such a comprehensive repository often leaves investigators struggling to work out whom to contact in another country, delaying investigations. Already participants are submitting their national research-misconduct policies and investigators' contact information to begin this central database. Other bodies such as the European Science Foundation and UNESCO (United Nations Educational, Scientific and Cultural Organization) are gathering similar information, and we plan to work with them to pool our resources and minimize duplication of effort. Which organization will maintain such information has still to be decided.

"Differences between national policies create challenges that must be resolved fairly and objectively."

We also suggested that the sorts of agreements — often known as 'memoranda of understanding' (MOUs) — signed at the start of international research collaborations should include a general statement of the value of research integrity and recommendations of what to do if an allegation is made. We have learned from experience that it is better to have a rubric for handling research misconduct before an incident occurs. Rash decisions and poor processes can result when people have to react quickly to determine whether a situation meets a definition of research misconduct or to sketch a plan of how to handle the matter on the fly. So we identified among other things the need to define misconduct, state who is in charge of investigations (and the procedures they will follow), detail how information will be shared during investigations, and specify how confidentiality will be assured.

We decided to broaden the membership of the committee, to better reflect the diversity of international research and its systems. We are inviting to future meetings representatives from other countries such as China, India, Russia, Argentina and Brazil, and institutions such as the Islamic World Academy of Sciences and the African Academy of Sciences, as well as other OECD countries such as Australia and Sweden. This will offer the committee a greater understanding and appreciation of each entity's legal and administrative system.

We charged several subgroups to further explore some of the issues discussed. One, for instance, will develop recommended frameworks and template agreements for international research collaborations. Another subgroup is defining relevant principles of investigations.

Next steps

These subgroups will report their preliminary results at the next expert-group meeting in Paris on 21–22 April 2008. Here we hope to reach agreement on the major issues and plan out the remaining work. By the end of 2008, the committee aims to publish a practical manual on how to approach misconduct cases in international research collaborations, and a directory of who to call in each country.

There are, however, thornier issues that the committee will not tackle. For instance research-misconduct prevention is not in our purview, neither is dealing with all that is involved in developing and ensuring responsible professional practices. Although important, we leave such topics to another expert group to explore seriously.

In a perfect world, the committee's work would never need to be implemented, and research misconduct would not be front-page news. Until then we must continue to focus on creating a practical framework for examining allegations and for holding researchers accountable for the integrity of their research.

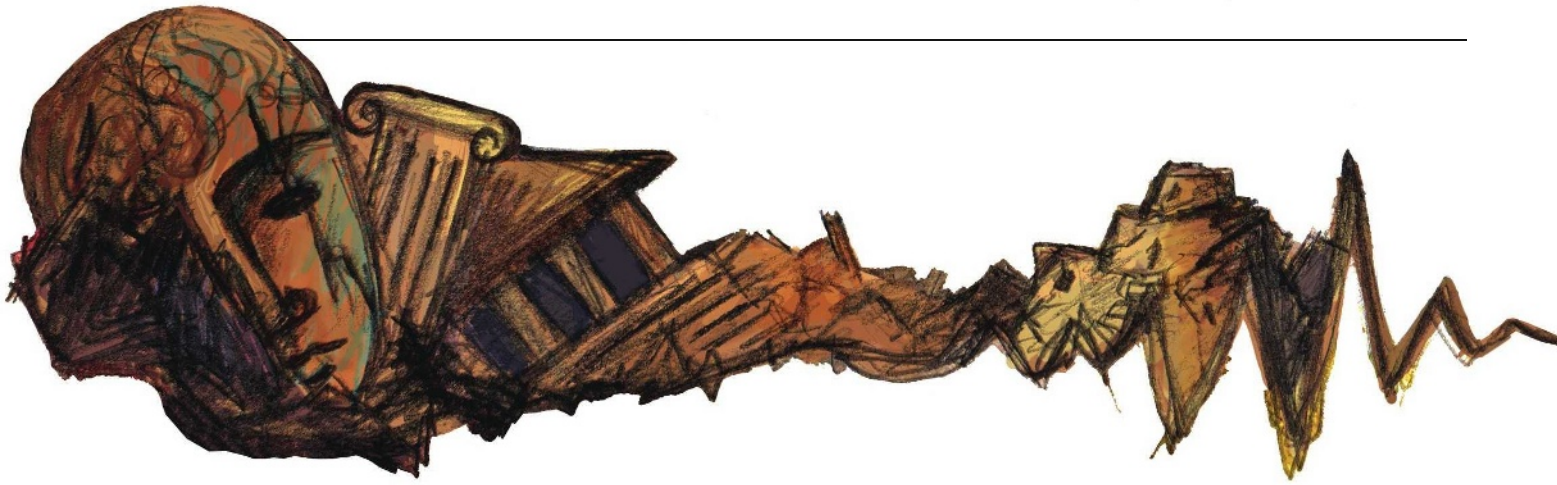
Such a task can be accomplished only with the input of the larger research community. Specifically, we are interested to learn of international misconduct-related documents that could inform the templates we hope to produce. We would also like to hear about any situations or challenges that scientists have encountered while conducting international research and if and how they were resolved. Researchers unfortunately find themselves in collaborative situations that go awry (see page 682) and they may be victimized by another who has conducted research misconduct. We therefore invite you to be a part of a global discussion on policing international misconduct at Nature Network. We look forward to your input. ■

Christine Boesz (representing the United States) is Inspector General of the National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230, USA.
Nigel Lloyd (representing Canada) is Executive Vice-President of the Natural Sciences and Engineering Research Council of Canada, 350 Albert St, Ottawa, Ontario K1A 1H5, Canada.

See Editorial, page 665.
Join the discussion at: http://tinyurl.com/2m8ntb

* The committee is an expert group of the Global Science Forum (GSF). For more information visit www.oecd.org/sti/gsf. † Australia, Germany, Poland, South Africa and Turkey are also part of the initiative but were unable to send representatives. ‡ The consensus paper was developed following a Workshop on Best Practices for Ensuring Scientific Integrity and Preventing Misconduct in Tokyo, held by the GSF and the Ministry of Education, Culture, Sports, Science and Technology of Japan. The consensus paper can be found at www.oecd.org/dataoecd/37/17/40188303.pdf.

SPRING BOOKS



Shaking the foundations of archaeology

Did earthquakes trigger the collapse of ancient civilizations?

Apocalypse: Earthquakes, Archaeology and the Wrath of God
by Amos Nur and Dawn Burgess
Princeton University Press: 2008. 304 pp. $26.95, £15.95

Andrew Robinson

At 5:04 p.m. on 17 October 1989, Stanford University geophysics professor Amos Nur was sitting in his office in California when it started to shake. His steel bookcases toppled and crushed his chair. Ducking under his desk, he somehow escaped injury. Nur had just experienced the Loma Prieta earthquake that devastated parts of the San Francisco Bay Area.

Nur has long been interested in archaeology. Growing up near Haifa in Israel, he walked behind his father and two mules as they ploughed the family's field. "I remember picking up shards of ancient glass," he writes, "pieces of simple mosaics and little squares of pink or white limestone exposed in the turned earth." On a few rare occasions, the plough struck a massive object: a rough block of stone about 1 metre tall and 25 centimetres square in cross-section. "Probably a Roman mile marker," they thought. Much later, Nur discovered that the field was on the *Via Maris*, the ancient Roman highway connecting the Mediterranean to the East.

A twin passion for seismology and archaeology drives Nur's deeply researched and compellingly written book, *Apocalypse,* co-authored with Dawn Burgess. In it he asks how earthquakes might be detected in the archaeological record, by analysing geological formations, faults, structural movement, human remains, the collapse of pillars and walls, and inscriptions. Nur wonders if earthquakes played a part in the collapse of ancient civilizations. Might they explain the enigmatic and quick disappearance of so many Bronze Age civilizations in the eastern Mediterranean during a mere 50 years around 1200 BC?

Most archaeologists today say that earthquakes have had little to do with historical demises. They prefer to attribute the collapse of civilizations to human agency: war, invasion, social oppression, environmental abuse and so on. The conventional explanation of the Bronze Age collapse involves maritime invasion by the mysterious Sea Peoples, whose identities have long eluded scholars. There are notable exceptions of academics who were sympathetic to the idea that earthquakes could crush civilizations — Arthur Evans at Knossos in Crete, Carl Blegen at the Turkish city of Troy and Claude Schaeffer, for instance — but the majority are sceptical.

Robert Drews took pains to quash any earthquake explanation in *The End of the Bronze Age: Changes in Warfare and the Catastrophe ca. 1200 BC* (Princeton University Press, 1993), as Nur is contemptuously aware. Jared Diamond made no mention of earthquakes or volcanic eruptions in *Collapse: How Societies Choose to Fail or Succeed* (Viking, 2005). If earthquakes really have had so great an influence, the sceptics ask, then where is the hard evidence?

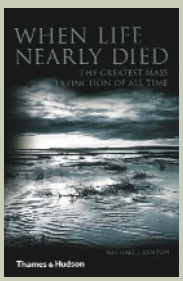
In reply, Nur and Burgess cite several powerful instances. The seismic destruction of the Portuguese capital of Lisbon in 1755, which provoked Voltaire to write *Candide*, shook the pillars of both religious faith and Enlightenment optimism. "By striking at a time when there was a particularly delicate balance of power between church and state, and between science and religion, the earthquake tipped the scales and changed society around the world," the authors argue.

In Venezuela, an earthquake in 1812

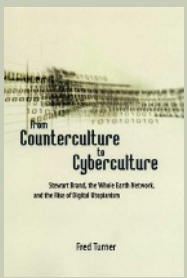
ILLUSTRATIONS BY C. ALLEN-FLETCHER

NEW IN PAPERBACK

When Life Nearly Died: The Greatest Mass Extinction of All Time
by Michael J. Benton (Thames & Hudson, $24.95)
The end-Permian extinction 250 million years ago destroyed 90% of Earth's life. Delving into geological history, Michael Benton investigates the science that documents the event, how its true scale was discovered, and discusses evidence for the possible cataclysm that caused it.



From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism
by Fred Turner (Univ. Chicago Press, $17)
By focusing on the life of ethical entrepreneur Stewart Brand and the influential Whole Earth Network, Fred Turner shows how alternative culture ideas in San Francisco, California, gave rise to the networked world.



precipitated the collapse of Simón Bolívar's republic by his own reckoning. This ultimately led to Bolívar freeing Colombia, Ecuador, Peru and Bolivia from Spanish rule.

In Japan, the Great Kanto earthquake of 1923, which reduced two-thirds of Tokyo to ashes, spawned political and racial turmoil that contributed to the rise of militarism and, ultimately, to the Pacific war. If the Tokyo area experiences another such earthquake in decades ahead as seismologists expect, its repercussions will surely make the global financial system tremble. Were it to strike at a time of economic depression, its effects might be globally catastrophic.

Apocalypse focuses mostly on the ancient world, with a distinct emphasis on biblical archaeology. It discusses earthquake evidence from the Middle East, including Jericho, Megiddo (Armageddon), Jerusalem and Qumran, the location of the 2,000-year-old Dead Sea Scrolls.

In the caves at Qumran, Nur has considerable field experience, which he deploys to illuminating effect. He was part of an expedition from Jerusalem's Hebrew University that excavated the rubble in the Cave of Letters, in the hope of finding a previously glimpsed skeleton and other evidence of habitation buried by the collapse of the roof in an ancient earthquake. Nur is convinced — a little like Howard Carter in the Valley of the Kings — that there remain sealed caves that were not looted by the Bedouin who first reported the scrolls' existence in 1947. "These places, undisturbed since their destruction by earthquakes, may provide the means to unravel the complicated and emotionally charged story of the Dead Sea Scrolls."

Apocalypse is a winning combination of cautious interdisciplinary investigation and interpretation, writing suitable for a general readership, and excellent illustrations (including a striking photograph of Nur's own crushed office chair). Although it will deliberately irritate many archaeologists, it should also provoke a serious reconsideration of the archaeological record. As with the evidence for human activity in climate change, the evidence for earthquakes in prehistorical change may be staring archaeologists in the face. ■

Andrew Robinson is a Visiting Fellow of Wolfson College, Cambridge, CB3 9BB, UK, and author of *Earthshock*.

Brave new bioethics

Life As It Is: Biology for the Public Sphere
by William F. Loomis
University of California Press: 2008.
272 pp. $24.95

Eugenie Scott

Science's task is to explain the natural world: what it is, how it works and why it is the way it is. Ethics is about the oughts and the shoulds. Most ethicists — religious and secular — agree that knowledge of the natural world helps us make better, or at least better-informed, ethical decisions. But, as David Hume, Thomas Henry Huxley and G. E. Moore have noted, a particular understanding of nature does not dictate a unique moral stance. For every Alexander Pope declaring "Whatever is, is right," there is a Rose Sayer (from the film *The African Queen*) retorting, "Nature ... is what we are put in this world to rise above!"

It is the complicated interplay of moral decisions and biological sciences that motivates cell biologist William F. Loomis. His brief book, *Life As It Is*, is a tour of the brave new biology relevant to such social issues as abortion, euthanasia, the use of embryonic stem cells, cloning, overpopulation and global warming. Loomis holds that scientific evidence should be taken into account when making socially important decisions. He provides a fascinating, if occasionally disjointed, survey of topics that bear on these decisions: the nature and evolution of life, and current scientific thought regarding consciousness, psychology and social behaviour.

Sometimes it is questionable whether the scientific aspects of a situation are most relevant to the ethical decision. Is it ethically permissible to destroy the surplus human embryos created for *in vitro* fertilization (IVF), for example? Loomis believes the answer should be shaped by a better understanding of the nature of cells.

Loomis emphasizes that at the cellular level life is cheap: at any given moment, billions of bacteria in our body are dying. A human zygote is merely a single cell, so shouldn't we think of it as such rather than the multicellular, functioning, conscious and precious baby into which it might develop? If a zygote is just a cell, and cells die regularly, then the answer to whether it is ethically permissible to destroy it is yes. But this argument comes after the ethical question of whether a zygote is just a cell, which is one that science cannot answer.

The ethical status of a human zygote or early-stage embryo turns on the issue of personhood. For those who believe in a soul, the moral standing of the zygote is largely unaffected by the nature of life at the cellular level. Belief in souls is a first principle, unlikely to be either proved or disproved by science.

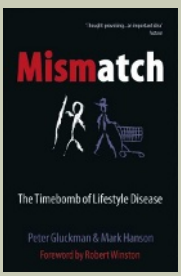
By contrast, as Loomis correctly notes, science may provoke a rethinking of religious dogma. Catholic theology holds that a soul is infused into a fertilized egg. So if an eight-celled embryo can be made to produce eight separate human beings, do they share a soul, or are seven new souls somehow generated?

This conundrum has led some Catholic theologians to contend that the soul is infused not at fertilization, but only when cells of the dividing organism lose their plasticity. Other theologians try to accommodate scientific facts about cells in other ways. Although scientific facts about the nature of a developing embryo may have profound consequences for Christian (or at least Catholic) thinking about souls — as a first principle, the concept of a soul is unlikely to be abandoned, and will be a factor in ethical decisions about many issues that biology touches on.

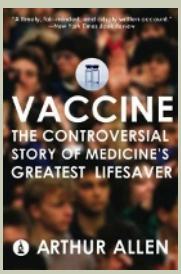
Policy-makers deciding between contending positions are ultimately forced to make political decisions, not scientific ones. Science — ideally, and in most cases — influences the thought of the proponents of the contending positions, and they in turn influence the policy-makers. But science is rarely the deciding factor. In many cases, such as the example of the human embryos in IVF, the contenders on both sides can agree on the science and disagree on the policy, owing to a disagreement about whether (and which) religious concerns are most relevant. And such disagreements are beyond the competence of science to adjudicate.

Refreshingly, Loomis's discussion of ethical issues roams beyond the comparatively narrow issues of abortion and euthanasia and the

Mismatch: The Timebomb of Lifestyle Disease
by Peter Gluckman and Mark Hanson (Oxford Univ. Press, £8.99)
The bodies we have now are the product of evolution. Peter Gluckman and Mark Hanson argue that they are mismatched to our needs in society today, and that this divide has increased the rate of lifestyle diseases such as diabetes and obesity.



Vaccine: The Controversial Story of Medicine's Greatest Lifesaver
by Arthur Allen (W. W. Norton, $17.95)
Journalist Arthur Allen investigates the history of vaccination, covering three centuries' worth of controversies. Reviewing the hardback edition, Michael Oldstone wrote that, "What becomes clear ... is that, when facts tangle with culture, culture often wins" (*Nature* **448,** 137; 2007).



like. He devotes a final important chapter to sustainability. In the face of pollution, global warming and population increase, how will it be possible to ensure an adequate supply of food, water and energy for all of Earth's people while maintaining respect for the well-being of other creatures? Loomis recommends a programme of voluntary population reduction, requiring both political leadership and a radical change of public opinion.

Loomis identifies the source of his title *Life As It Is* — his wife apparently — but not its significance. The idea that a realistic understanding of biology will usher in a paradise of ethical correctness is naive: the panoply of extra-scientific considerations that influence ethical decision-makings cannot be ignored or minimized. A weakness of Loomis's book is his comparative neglect of such considerations. But if his intention is less ambitious, namely that a realistic appreciation of biology ought to inform ethical decision-making, then that is incontrovertible. ■

Eugenie Scott is executive director of the National Center for Science Education, Oakland, California 94609, USA, and author of *Evolution vs Creationism: An Introduction*.

Engines of life

Energy in Nature and Society
by Vaclav Smil
MIT Press: 2008. 512 pp. $70, £45.95

Tim Lenton

The explosion of myriad life-forms throughout Earth's history has been fuelled by their ability to collect and process increasing amounts of energy. Thus organisms have become ever more complex, culminating in humans and our technology. In *Energy in Nature and Society*, Vaclav Smil describes in quantitative detail the evolutionary and technical innovations responsible, from photosynthesis and respiration to solar cells and steam turbines.

The first living things probably accessed chemical energy by breaking down large molecules into smaller ones. By 3.7 billion years ago, the first photosynthesizers evolved. These organisms could capture energy from sunlight and use it to split simple molecules and liberate electrons, which they used to make sugars from carbon dioxide. Their ingredients probably included electron donors — hydrogen (H_2, then H_2S) and later iron (Fe^{2+}) — that were in limited geological supply. This restricted global productivity to at most a tenth that of the modern marine biosphere.

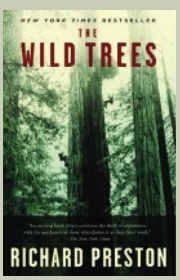
Next came the greatest energetic innovation in the history of the planet: oxygenic photosynthesis, the ability to capture enough energy from sunlight to split water, thus liberating oxygen gas. This evolved in cyanobacteria more than 2.7 billion years ago. Initially, oxygen production was confined to microbial mats and sunlit surface waters; 2.4 billion years ago it rose in the 'great oxidation' of the atmosphere.

When oxygen reacts with organic matter during aerobic respiration, an order of magnitude more energy is liberated than was available for earlier anaerobic respiratory pathways. Ultimately, this source of power allowed the evolution of larger and more mobile organisms, including humans.

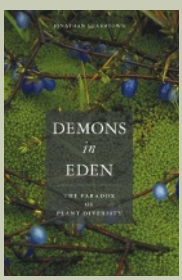
Smil is brief on the history of the biosphere. He gives a fascinating assessment of its present energetic state, quantifying global energy capture in photosynthesis and the uses that all organisms, including humans, put it to. Then he shifts the focus to human technological innovations that have progressively increased the capture and conversion of energy into forms that are useful to us. *Energy in Nature and Society* tells this story wonderfully, from hunter-gathering to traditional agriculture, the shift from human to animal power, the invention and refinement of water wheels and windmills, improvements in roads and ship design, and to charcoal production and its use in metallurgy. The fossil-fuel age takes off with exponential global increases in coal, then oil, then gas extraction and consumption.

In a feat unprecedented for a single animal species, humanity's total energy use has now exceeded that of the entire ancient biosphere before oxygenic photosynthesis, reaching about a tenth of the energy processed by today's biosphere. Almost half of the world's total primary energy supply is consumed by the rich

The Wild Trees
by Richard Preston (Random House, $16)
A dramatization of the lives of Californian botanists, Richard Preston reveals the hidden world of the coastal redwood trees, the tallest organisms the planet has sustained. Science, adventure and a passion for trees are combined.



Demons in Eden: The Paradox of Plant Diversity
by Jonathan Silvertown (Univ. Chicago Press, $16)
How is plant biodiversity maintained and why is the world not overgrown by aggressive weeds? Jonathan Silvertown explores the dynamics of the plant world and suggests that "tasting the fruit of evolutionary knowledge may provide us with a ticket for readmission to the Garden of Eden", according to Peter Moore's review (*Nature* **438,** 27; 2005).



G8 nations, despite their having only 12% of the world's population. The poorest quarter of humanity consumes less than 3%. For them, even modest increases in per capita energy consumption significantly reduce infant mortality and increase life expectancy. Above about 60 gigajoules per capita (the amount used, for example, by citizens of the French city of Lyon in 1960) these benefits level off, indicating that profligate energy use bears little on quality of life. Consequently, Smil advocates that rich people should reduce their energy consumption to allow poorer people to increase theirs.

Today's high-technology societies mostly rely on fossil fuels. These concentrated reserves of stored ancient sunlight — the remnants of past organisms — are finite, and the products of their combustion have undesirable consequences, from respiratory problems to climate change. Smil argues that we should stop the seemingly endless growth of energy consumption while we switch to cleaner and more sustainable sources of power.

In the long term, Smil's solution is solar power, because the total supply of sunlight at Earth's surface exceeds current global fossil-fuel consumption by more than a thousand times. Until then, he supports a careful transition away from fossil fuels and points out that carbon capture and storage have limited capacity. He dismisses most renewable energy sources because their power densities are too low to supply the needs of the present global population, let alone future ones. This includes first-generation biofuels, where their poor energy returns could ultimately mean feeding cars in place of people. In *Energy in Nature and Society*, Smil contends that power from nuclear fission would become limited by uranium supply until a viable commercial fast-breeder reactor is available, and fusion power is too far from commercial deployment. All of which implies a difficult transition period involving substantial carbon dioxide emissions and climate change.

If we are to have a long and happy future on this planet, we need to follow life's example and find more efficient ways of extracting free energy from sunlight. But energy isn't everything. The successful major transitions between past biospheres also required increases in material recycling, because the organisms capturing energy built their bodies from elements whose supplies were restricted. In contrast, to maintain our present high level of energy transformation in society, we increase the inputs of many elements, mostly by mining them from finite reserves in Earth's crust. We thus perturb global biogeochemical cycles. A wiser strategy in the long term would be to increase the recycling of materials that accompany the capture and use of energy.

I would like to encourage Smil to strengthen this link between energy processing and material cycling in the next edition of his book, and to address whether the combination of solar power and recycling might allow energy transformation by humans and the biosphere to grow again. For now, the energy required to read this comprehensive and scholarly tome is extremely well spent. ■

Tim Lenton is professor of Earth system science in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Biology from the bottom up

What Is Life? Investigating the Nature of Life in the Age of Synthetic Biology
by Ed Regis
Farrar, Straus & Giroux: 2008. 208 pp. $22

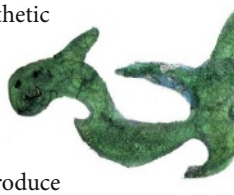
Steven Benner

Book titles should display ambition, and Ed Regis' latest certainly does that. Implicit is progress between two areas of biology. *What Is Life?* recalls Erwin Schrödinger's famous book of the same name that encouraged many physicists to begin working in molecular biology in the 1940s; synthetic biology is the fast-moving area today.

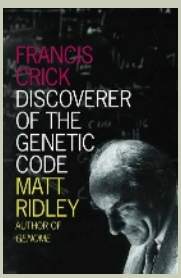
The term synthetic biology was coined in 1974 by Waclaw Szybalski to describe the modification of organisms by adding and subtracting genes. In those days it was known as 'genetic engineering' or 'recombinant DNA technology'. By altering the genes, the organisms act in new ways.

At the time, Szybalski's synthetic biology prompted fear. The city of Cambridge, Massachusetts, banned genetic engineering entirely. A conference was convened in Asilomar, California, to decide how to manage the new ability to create artificial organisms.

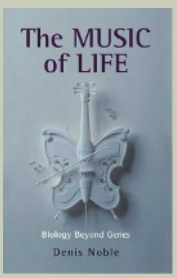
Three decades of experience have shown the risks to be negligible but the rewards enormous. Today, the field of synthetic biology is expanding, spawning new university departments, such as the one that hosts Jay Kiesling's laboratory at the University of California at Berkeley in which bacteria are created to produce



Francis Crick: Discoverer of the Genetic Code
by Matt Ridley (Harper Perennial, £7.99)
The story of Francis Crick extends beyond the discovery of DNA. Matt Ridley's biography details how he came to study biology, sets in context his controversial ideas and gives a glimpse into the character of one of the most famous scientists of the twentieth century.



The Music of Life: Biology Beyond Genes
by Denis Noble (Oxford Univ. Press, £7.99)
Instead of taking a blinkered view based on genes and genomes, Denis Noble argues, we must realize that life is a process. To see its workings, we should look at interactions on every level — genes, cells, the body, systems and the environment.



pharmaceutical intermediates. Craig Venter, a driver of innovation in contemporary genomics, and whose personal genome can be found on the Internet, is going further by proposing reorganization of the natural parts of natural genomes. Some of these restructured microbes are so scrambled that they deserve to be viewed as new species.

The remit of synthetic biology has widened as other researchers have adopted the label. In 2000, Eric Kool of Stanford University, California, used it to describe the construction by chemists of unnatural molecules that can operate within natural living systems. To Drew Endy and others at the Massachusetts Institute of Technology, it means the process of creating, mostly by modifying existing biomolecules, units that can serve as interchangeable parts in larger assemblies. Stephen Wolfram and others view "artificial life" as a computer program that yields output behaviour that is analogous to the behaviour of living systems.

What is Life? captures these differing perspectives well. As expected from a science writer with Regis' record, the book is an easily readable review of the development of contemporary biology, including the first-generation model for DNA structure, the foundation of metabolism, and the elucidation of the genetic code. Furthermore, it captures interactions between scientists who approach synthetic biology differently, providing a brief and entertaining glimpse into the competitive aspects of modern science. For example, one experimenter (Norman Packard of Protolife, based in Venice, Italy), trying to get a real cell made out of real chemicals to work in a real laboratory, sets these activities above trying to write computer programs that simulate parts of biological chemistry. Another (Francis Collins, who heads the National Human Genome Research Institute in Bethesda, Maryland) is quoted asking, in essence: what's new? Isn't this just the 30-year-old field of genetic engineering sporting a catchier trademark?

There is one disappointment. The book only incompletely conveys why efforts to rebuild life from the ground up ('synthesis') offer new avenues for discovery that those dissecting life from the top down ('analysis') do not.

The analytical approach to biology was born a few centuries ago, when those wishing to answer the question, 'what is life?', realized that observation alone was insufficient. Their investigations began by killing some unfortunate organism. After dissecting the spilled guts, tissues were named, maps were drawn and parts were catalogued. Much was learned; much of it practical. But the essence of 'life' did not emerge. With the invention of microscopes, the dissection went further, to cells. This time a new theory (cell theory) did emerge. As Regis' book emphasizes, cells are even today viewed as a defining attribute of life.

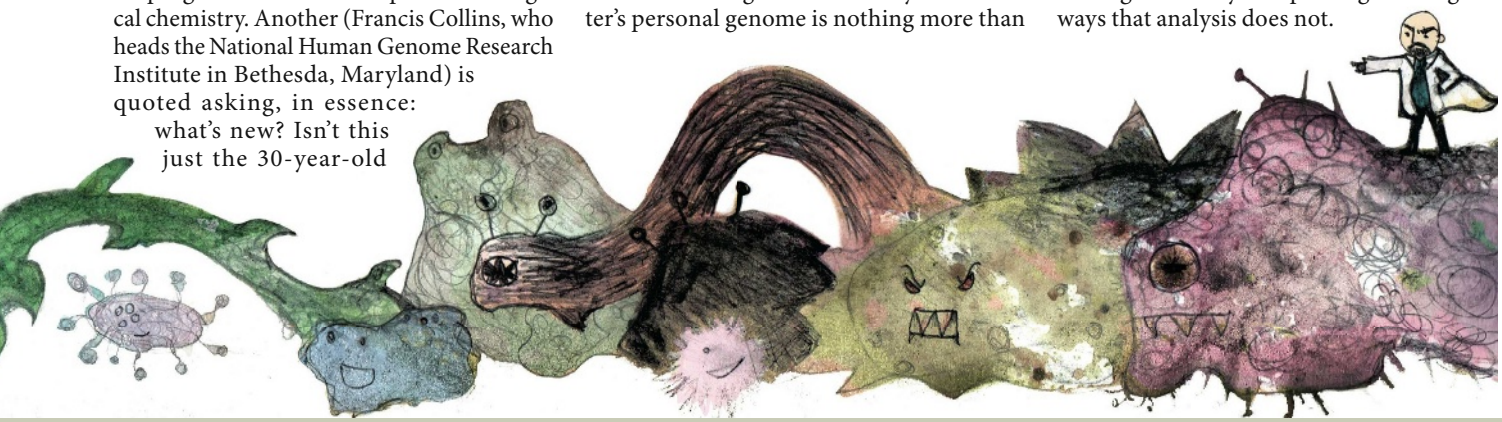
The so-called 'age of biology' came not from biologists but from chemists, who carried the dissection of living matter further. Karl William Scheele, in the late eighteenth century, crystallized the first organic molecule (barium lactate) from sour milk, and realized that the molecular parts of living organisms could be analysed. This led to structure theory, which holds that the arrangement of atoms in constituent molecules determines the behaviour of all matter. Biological chemists spent the next 150 years figuring out atomic arrangements in every biomolecule they could get their hands on, even DNA.

An unbroken line runs from Scheele to the human genome. It involves great technological innovation, but no conceptual innovation that can be thought revolutionary. Even Venter's personal genome is nothing more than a map of how its atoms of carbon, oxygen, hydrogen, nitrogen and phosphorus are arranged.

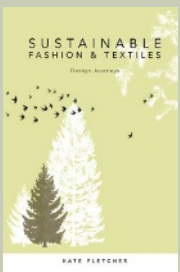
Even if the analytical strategy applied to biology is ever completed, biology will remain hollow. Living systems cannot be explained solely as a series of molecular structures, even when their interactions are described mathematically (as attempted in systems biology). Reflecting this, microbiologist Carl Woese wrote that the "strange claim by some of the world's leading molecular biologists that the human genome is the holy grail of biology is a stunning example of a biology that has no genuine guiding vision".

Synthesis offers a different strategy. The deliberate creation of new forms of matter from the bottom up, rather than the top down, gives us new ways to test nature. Chemists today use synthesis routinely. Having benefited from being first to gain the tools, they tested structure theory by building molecules with structures designed to target predictions of the relationship between molecular structure and behaviour. In a virtuous circle, they simultaneously built up their molecular toolkit and improved structure theory, further empowering synthesis. Chemists know that if one truly understands a phenomenon, one should be able to synthesize another, different system that generates that phenomenon.

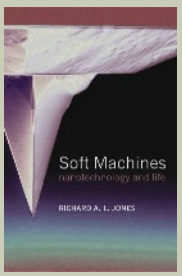
Because building something requires a deep understanding of its parts, synthesis also stops scientists from fooling themselves. Data are rarely collected neutrally during analyses by researchers, who may discard some, believing the data to be wrong if they do not meet their expectations. Synthesis helps manage this problem. Failures in understanding mean that the synthesis fails, forcing discovery and paradigm change in ways that analysis does not.



Sustainable Fashion and Textiles: Design Journeys
by Kate Fletcher (Earthscan, $48.95, £24.95)
Fashion is ephemeral by nature, but Kate Fletcher describes how clothing manufacture could be turned into a sustainable industry. Her detailed book assesses systems as well as products, examining possible solutions from raw material to final design.



Soft Machines: Nanotechnology and Life
by Richard A. L. Jones (Oxford Univ. Press, £9.99)
The principles of nanotechnology may have more in common with biology than engineering, argues Richard Jones. He describes the science of the minuscule, and explains how the demands of working at this scale may shape nanotechnology into something more organic.



Now that genetic engineering is available, biologists are benefiting. By attempting to create synthetic genetic systems, we will learn more about how natural genetic systems work; by attempting to create synthetic metabolisms, we learn about how natural metabolisms work; by attempting to create synthetic regulatory circuits, we learn about how natural regulatory circuits work.

Will we ever understand what life is? Just as with Schrödinger's book, Regis' text will not be the last word. It is, however, a good place for a lay reader to start, one who welcomes the ambition of its title. ■

Steven Benner is a distinguished fellow at the Foundation for Applied Molecular Evolution and the Westheimer Institute for Science and Technology, Gainesville, Florida 32601, USA, and co-author of *The Limits of Organic Life in Planetary Systems*.

Bottling the nuclear demon

Why Nuclear Disarmament Matters
by Hans Blix
MIT Press: 2008. 97 pp. $14.95, £9.95

On Nuclear Terrorism
by Michael Levi
Harvard University Press: 2007. 210 pp. $24.95, £16.95

Martin van Creveld

Right or wrong, nuclear proliferation is much in the news. These two works tackle the problem head on. The first is a somewhat emotional call to prevent proliferation from proceeding further and, if possible, to reverse it. The second deals with some of the problems to which it may give rise.

To the readers of *Nature*, as well as anybody who is familiar with the origins of the second Iraq War, Hans Blix needs no introduction. In 2002–04, the elderly, genial Swedish diplomat, former foreign minister and former head of the International Atomic Energy Agency, found himself at the head of the UN Monitoring, Verification and Inspection Commission (UNMOVIC), charged with finding weapons of mass destruction in Iraq. Announcing that he had failed to discover any and that they almost certainly did not exist, he had to confront the full wrath of the Bush administration — a story that throws an ugly light on that administration while showing Blix himself in a very positive one.

A reader looking at the title of Blix's new work might be forgiven for thinking that he provides a retrospective account of several decades-worth of effort to put the nuclear demon back into the bottle from which it had escaped. He does nothing of the kind. First, contrary to his promise, he does not focus on nuclear weapons alone but widens the discussion — and, to my mind, weakens his case — by including chemical and biological ones too. Second, part of the book has little to do with nuclear disarmament but constitutes a polemic against the Bush administration's attempt to develop a national strategy based on pre-emption. Given how slim the volume is — in reality, it is just a brochure — this leaves little room for a serious discussion of nuclear disarmament, why it matters and what steps towards its realization should be taken next.

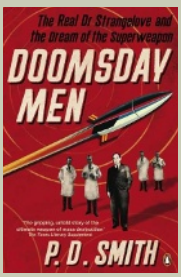
Furthermore, Blix is what the Germans call a *Gutmensch* — a do-gooder. From his perch in lucky Sweden, he observes the world with a benign smile and gently seeks to teach other benighted people how to improve it. One may certainly respect this position. On the other hand, it is neither the only one possible nor necessarily the one best suited for dealing with future nuclear-armed Adolf Hitlers, Joseph Stalins, and, yes, Saddam Husseins as well.

To put it bluntly, Blix is no strategist. As he himself comes very close to saying at the beginning of this book, he thinks in terms of morality and well-being, not of power. Not once does he mention deterrence. In other words, the fact that, had it not been for nuclear-weapons proliferation, World War Three might very well have broken out long ago and perhaps obliterated both his native country and himself.

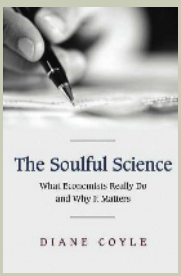
A much younger man than Blix, Michael Levi is almost unknown. The cover of his book merely says that he is "Fellow for Science and Technology at the Council on Foreign Relations, New York". From an Internet search, we learn that he is an academic who has worked here and there and published this and that. Yet anyone who reads his work cannot but be impressed by his deep understanding of nuclear terrorism and the possibilities of dealing with it.

Levi's work is written in a calm, unemotional and somewhat dry manner. Those looking for hair-raising accounts of how

Doomsday Men: The Real Dr Strangelove and the Dream of the Superweapon
by P. D. Smith (Allen Lane, £8.99)
In the 1950s, humans became capable of destroying life. Smith describes the first weapons of mass destruction and how the doomsday bomb became a symbol of science's destructive power. "The book is as much a history of modern science as of modern weaponry", wrote Gregg Herken (*Nature* **448,** 868; 2007).



The Soulful Science: What Economists Really Do and Why It Matters
by Diane Coyle (Princeton Univ. Press, $19.95)
Economics is not a dry science but a human one, says Coyle. By incorporating psychology, evolution and complexity, economists are in the best position to model human society. Frances Cairncross wrote: "To understand how the big ideas of the past half-century fit together ... read this book" (*Nature* **447,** 1057; 2007).



bad terrorists almost succeeded in blowing up the world but were prevented from doing so at the last moment by good policemen and other intelligence personnel will be disappointed. Indeed it is one of Levi's strong points that, from beginning to end, he does not engage in any kind of mystery-solving or moralizing. What he does provide is a step-by-step account of the many possibilities of using nuclear bombs and materials for terrorist ends and of at least some of the things that can be done to prevent those possibilities from being realized.

To cut a long story short, Levi sees many serious obstacles standing in the way of would-be nuclear terrorists. They start with the near impossibility of producing (as distinct from purchasing or stealing) high-grade uranium and plutonium; they end with the fact that any nuclear weapon a terrorist may build will almost certainly be crude, very heavy (and thus difficult to conceal and to move) and, quite possibly, unreliable as well. Other possibilities such as a 'dirty' bomb also exist, but compared with a full-scale bomb they are relatively harmless.

That is the good news. The bad news, Levi tells us, is that nuclear terrorism is not impossible in principle. Nor will any set of precautions, however well considered, provide 100% security against it.

Levi's argument may be boiled down to five propositions. First, policy-makers should make "the strongest possible efforts to improve controls over nuclear weapons and materials and prevent their further spread". Second, strategic assessments of the threat should avoid the mistake of concentrating solely on the worst possible scenario. Instead, they should also consider the most likely ones and, to encourage thought and cooperation, as far as possible those assessments should be made public. Third, if nuclear terrorism is to be prevented, it is necessary to set up a global system. The more governments cooperate and the more numerous and varied the obstacles they confront terrorists with, the greater the chances of success. Fourth, even imperfect countermeasures are better than none, because they may well sound the alarm.

Fifth and perhaps most controversial, it is important that thought be devoted, and steps taken to reduce damage from nuclear terrorist acts after they have taken place — given that, in Levi's view, some such steps should be able to dramatically reduce casualties.

Levi's measured text does not rely heavily on mathematics, and he has hidden the few equations he uses in the appendix. Still, here and there the book, with its flat tone and relatively few verbs, is not altogether easy to understand. Moreover, as he notes, there are certain things that might be useful to terrorists and that, accordingly, he cannot say. These are minor shortcomings if, indeed, they are shortcomings at all. Although the volume will not enjoy a wide readership, for anyone with a serious interest in the possibilities of nuclear terrorism and how to prevent it, *On Nuclear Terrorism* is a must.

Both authors assume that nuclear proliferation is bad. In Blix's case, this is because it just is so; in Levi's, because it increases the chances of nuclear terrorism. Neither gives a thought to the possibility that proliferation and the balance of terror have actually been excellent things — perhaps the best that ever happened to mankind. The difference between them is that, given Levi's theme, his ignoring this entire question is apt. In the case of Blix, it is certainly not. ■

Martin van Creveld is professor emeritus of the Hebrew University, Jerusalem 91905, Israel. His most recent book is *The Changing Face of War: Lessons of Combat, From the Marne to Iraq*.

How music speaks to us

Music, Language, and the Brain
By Aniruddh D. Patel
Oxford University Press: 2008. 528 pp. $59.95

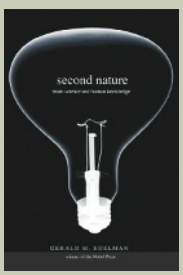
David Poeppel and Elika Bergelson

This book is an intellectual tour de force, raising many more issues than recent popular works by, for example, Oliver Sacks and Daniel Levitin. Not one for the bus, beach or bathtub, *Music, Language, and the Brain* requires focused engagement, but its rewards are rich. Aniruddh Patel offers a thorough analysis of music cognition and its relation to language, and outlines an ambitious and innovative research programme that deepens our understanding of cognition in general.

Music and speech share basic sound elements, and Patel starts by highlighting the similarities and differences between how auditory signals work. The book then delves into five topics: rhythm, melody, syntax, meaning and evolution. Each topic is examined within the context of music and language, to see how key cognitive processes overlap or diverge. By evaluating the latest empirical evidence, the author proposes further studies to test or extend previous results — experimentation, he says, is crucial in moving this field forward. Clearly, Patel has particular theories that he favours, but he describes fairly the ideas of others. The book is admirably clear in stating what has been done, and what needs to be done.

The belief that there are fundamental similarities in the processing of music and language is largely intuitive and worth testing. Both have been argued to be unique to humans. The book

Second Nature: Brain Science and Human Knowledge
by Gerald M. Edelman (Yale Univ. Press, $13, £8.99)
Nobel prizewinner Gerald Edelman offers a new theory of knowledge based on brain science. He shows how advances in neuroscience and physiology have led to a greater understanding of the brain, consciousness and creativity.



Skin: A Natural History
by Nina G. Jablonski (Univ. California Press, £9.95)
Nina Jablonski explores skin's many purposes, such as its role in touch and emotional display, and explains them as the result of billions of years of evolutionary compromise. "*Skin* is not just about biology, but also the way we live," wrote John Galloway (*Nature* **445,** 367–368; 2007).

emphasizes the particulate nature of music and language — both assembled from discrete elements — and suggests that these two domains may share a set of brain structures. By contrast, studies of brain lesion data (from patients with deficits that follow specific brain damage) and brain imaging results are also consistent with a view that music and language processing are, at least in part, segregated.

Patel thinks that there are more general, perhaps computational, links between the two. For example, when discussing rhythm, he proposes that the processing of non-periodic signals is similar in both music and speech. In the section on syntax, he argues that the brain uses similar neural resources to integrate the hierarchical organization of music and language. When discussing evolutionary and developmental similarities, the notion of 'beat-based rhythm processing' emerges as a crucial feature that may underlie music and speech.

Patel's perspective is laudably cross-linguistic and multicultural, citing extensive work from non-Indo-European languages and non-Western-based musical systems. On the website accompanying the book (http://tinyurl.com/2z2cve), Patel provides stimulating sound and video examples that clarify the phenomena described in each chapter.

Music, Language, and the Brain is much more than a textbook by one of the field's most influential practitioners. Each chapter can serve as a stand-alone monograph, and can be read at many levels. There is enough clarity for the general reader to follow the lines of argument, while the specialized reader will discover Patel's sophisticated and well-researched positions. Ideal for students of music cognition and language, the book outlines numerous experiments and hypotheses — many unusual — that draw together psychology and neurobiology.

If one can criticize anything, it is that Patel's discussion of the neurobiological foundations of auditory cognition is less nuanced and inspirational than his treatment of behavioural research. That said, we know little about the neuronal bases of complex psychological phenomena. Our understanding of auditory cognition is still mostly informed by behaviourally based psychological research, and in that domain, Patel's discussion is second to none.

In this definitive analysis of music cognition and its relationship to language, Patel gives us a work of exceptional scholarship and clarity. Much needed, it elevates the discussion to a level that these exciting areas merit. ■

David Poeppel is a professor of linguistics and biology and Elika Bergelson is a Baggett research fellow at the Linguistics Department, University of Maryland, College Park, Maryland 20742-7505, USA.

Catching a ride on sunshine

Solar Sails: A Novel Approach to Interplanetary Travel
by Giovanni Vulpetti, Les Johnson and Gregory L. Matloff
Springer: 2008. 250 pp. £16.50

Stuart Clark

Conceptually simple and romantic, solar sailing is an enchanting technological solution for space exploration. When a large reflective sail is unfurled in space, photons of sunlight collide with the sail fabric, imparting pressure and causing the sail to move. Such photons are not the electrically charged particles that constantly flow from the Sun to create the solar winds, they are the actual sunlight itself. The angle of the sail to the Sun and its direction of travel determine whether a propelled craft speeds up or slows down, just as a yacht changes course on the sea.

Solar Sails: A Novel Approach to Interplanetary Travel is the latest book to explore this topic, one that has been tackled only a handful of times in the past 20 years. Aimed at undergraduates, the book convincingly captures the history of ideas about solar sails, their current state of play and their future promise.

Moving according to the constant interplay of gravity and the pressure of sunlight, spacecraft pushed by solar sails are highly manoeuvrable. They can skate along unusual interplanetary trajectories that traditional point-and-shoot rocket-propelled craft would find difficult, if not impossible, to navigate. In the flexibility stakes, the only current competition is from the newly tested but expensive ion-drive engine that powers the SMART-1 Moon mapper built by the European Space Agency (ESA) and NASA's Deep Space One asteroid probe. These propulsion modules run by expelling charged particles, or ions, and can operate using less fuel than standard chemical engines; however, they are technologically trickier and thus expensive to build.

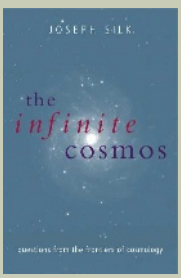
The idea that sunlight exerts pressure has been around for more than a century, since physicist James Clerk Maxwell proposed it in the 1860s. In the 1970s, metre-long solar sail fins — rather like the fins on a 1950s American car — were attached to the Mariner 10 Mercury space probe to adjust its alignment. Today, some satellites are steered with small sail vanes, a technology patented by the aerospace company EADS Astrium. The extra force of sunlight is a hindrance when fine control of movement is required, as with the next generation of formation-flying spacecraft in ESA's proposed Darwin interferometry mission to search for life on extrasolar planets. Such vessels must instead be designed to minimize displacements or, at least, to all suffer the forces equally.

Despite the opportunities, solar sails have yet to be used for propulsion in space. The pressure of sunlight is so slight that a vast sail area would be needed to carry a worthwhile payload of instruments through space. Deploying such a sheet presents an equally vast challenge, and has remained the solar sailor's Achilles' heel.

With useful sails being many tens to hundreds of metres long, these mighty structures must be packed into the equivalent of a suitcase for launching and then faultlessly unfurled once in space. If the sail snags, tears or fails to deploy, the mission is over. This risk deters many potential users; according to one project scientist at the ESA: "Why jeopardize your science by relying on an untested technology?"

Scientists and space agencies have, until recently, been resistant to solar sailing. This negative attitude was reinforced by the failure of the Planetary Society's Cosmos-1 sail, launched atop a converted Russian intercontinental ballistic missile on 21 June 2005 from a submarine in the Barents Sea north of Russia. The upper-stage rocket motor failed,

The Infinite Cosmos: Questions from the Frontiers of Cosmology
by Joseph Silk (Oxford Univ. Press, £9.99)
Summarizing the latest thinking on the Universe and its fate, Joseph Silk muses on the scientific discovery process and the history of ideas about the cosmos. "Black holes, galaxy formation, dark matter, time travel, string theory and the cosmic microwave background all get a mention" (Peter Coles, *Nature* **441,** 285; 2006).



King of Infinite Space: Donald Coxeter, the Man Who Saved Geometry
by Siobhan Roberts (Profile Books, £14.99)
Donald Coxeter helped to bring geometry back into the mainstream at a time when it was unpopular even with mathematicians. Siobhan Roberts' biography describes his personal and professional life and shows how his impact can be felt in architecture, cosmology, crystallography, immunology and more.

dooming the mission to failure before the sail mechanism could even be tested. Although the test was inconclusive, the perceived lack of success reflected badly on the solar-sail initiative itself.

Now the tide is beginning to turn. Ground-based tests in Europe and the United States have successfully deployed sails of about 20 square metres thanks to improvements in sail-opening mechanisms. The German Aerospace Centre has used plastic booms reinforced with carbon fibre, and NASA has used inflatable booms that harden when exposed to the coldness of space. Even more impressively, the Japanese space agency JAXA has carried out two successful sub-orbital deployment tests. Made of reflective films 7.5 micrometres thick and some 10 metres in diameter, the sails were flown to an altitude of 122 kilometres, where one opened up like a clover-leaf, the other like a fan. JAXA followed this up two years later in 2006, with a successful 20-metre-wide sail deployment from a balloon at an altitude of 35 kilometres.

Some space missions can be performed efficiently only with solar sails. Placing a satellite in a polar orbit around the Sun using a rocket requires a large expenditure of energy, and hence fuel. A craft propelled by a solar sail would take only five years to fly there from Earth but would require a huge sail area of 25,600 square metres (160 metres by 160 metres: larger than five American football pitches side by side) to provide the necessary thrust. Because sunlight holds the sail in space, it can be angled so it hovers like a kite over the poles of a planet, making solar-sail craft ideal anchors for communications and remote-sensing satellites.

Of course, there are limitations. Solar sails lose their power and manoeuvrability when they are far from the Sun, out beyond Jupiter. They are also unable to assume low orbits around planets with atmospheres because the sails are susceptible to drag.

Suitable for aerospace students and keen enthusiasts alike, this book may one day inspire some of them to build a solar-sail-powered vessel. Although there is still a long way to go, this useful volume will help speed up that day. ■

Stuart Clark is a visiting fellow at the Centre for Astrophysics Research , University of Hertfordshire, Hatfield, AL10 9AB, UK. He is author of *The Sun Kings*.

Imaging the unseen

Six Stories from the End of Representation: Images in Painting, Photography, Astronomy, Microscopy, Particle Physics, and Quantum Mechanics, 1980–2000
by James Elkins
Stanford University Press: 2008.
320 pp. $65

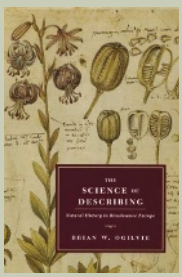
Felice Frankel

In *Six Stories from the End of Representation*, James Elkins makes a brave and laudable attempt to address in parallel the communication of ideas in both the sciences and the humanities.

"I believe that the clearest, most fruitful response to the abyss between the humanities and the sciences is to set out the disciplines, in detail, side by side, and let them tell their stories in their own languages. As far as I can see, that is the only way to produce a book that can be read by scientists and humanists without the creeping feeling that their disciplines are better explained — or explained away — by someone who does not really understand them," he writes.

Elkins' approach struck me as exciting. A respected and prolific scholar, he promised a view of scientific images distinct from that taken by other art historians. I was eager to see the images to which he was going to introduce us, and how they were to "tell their own stories". After all, Elkins declared that "the images came first" in preparing the book, and that the images "are among the best that have been made in the last several decades".

Science of Describing: Natural History in Renaissance Europe
by Brian W. Ogilvie (Univ. Chicago Press, $27, £14)
In the mid-sixteenth century, naturalists developed tools for observing and describing nature, enabling them to assess and share their findings with others. Interpreting this change over four generations, Brian Ogilvie "has written the story of how science constantly reinvents itself, seen through the lens of the pre-linnaeans", wrote Sandra Knapp (*Nature* **442,** 871; 2006).



Digital Art (Revised Edition)
by Christiane Paul (Thames & Hudson, £8.95)
The digital format offers new possibilities for artistic expression. In an updated volume, Christiane Paul addresses how viewers interact with such works, and explores links with artificial life and intelligence, activism and networks, as well as the collection, presentation and preservation of digital art.

So I can imagine Elkins's horror (my response, were I in his shoes) to discover the publisher's plan to illustrate this book of images by means of the insertion of a small collection of colour plates smack in the middle of the edition, forcing the reader to hunt through the text for reference to them, with the remaining majority of the images placed within the narrative but reproduced in grainy black and white on cheap paper. Surely publishing a book about seeing images warrants allowing the reader to see the images?

Beyond the layout, Elkins's approach offers a new way of making us look more closely at how we depict objects. The book is spread across six chapters, and includes more than 100 visual representations drawn from art and science created between 1980 and 2000. The art collection includes photography, sculpture and painting: works by Agnes Martin, Gerhard Richter, Sol LeWit and Jasper Johns, among others. The science collection is also broad, ranging from representations of the very small (particle physics) to the very large (astronomy), and including TEMs (transmission electron micrographs) of influenza virus, STMs (images from scanning tunnelling microscopes) of gold surfaces, Feynman diagrams of particle interactions and Hubble Space Telescope images of the galaxies.

Elkins informs us that these images are "objects that literally don't exist ... abstractions of abstractions, feeble symbols of objects that have no reality of their own". Leaving the philosophy of existence to other experts, I will not delve here into an argument of how many of the book's science images are representations of objects that do have a reality of their own. It is true that all images in science are representations of data of some sort, but I'd like to think those data derive from evidence of existence, even if transitory.

In telling his *Six Stories from the End of Representation*, Elkins stops short of making explicit links between images. He uses what he calls a "non-causal narrative", where "it is the reader's task to decide what threads might tie the images together and just how tightly they should be pulled". His juxtaposition of the blurred-looking astronomical photograph of the star Wolf-Rayet 104 and Edward Ruscha's intentionally blurred acrylic, *F House*, is used as an example of an "illusory coherence brought on by certain habits of seeing" — blurring, I suppose in this case.

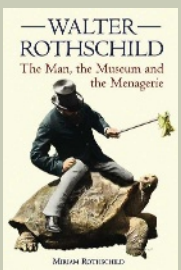
For Elkins, "it became clear that the images were the results of investigation into the limits of representation. Few of them contained sharp or well-defined objects, and in most it was difficult to understand what was being taken as adequate representation. They seemed to have a common theme: they were images that did not simply depict objects, but demonstrated how some objects resist depiction."

There is much I might challenge in Elkins's attempt to get the two cultures to talk to each other — first and foremost is his selection of "the most important" images. His choices are, after all, personal, and the book seems to be more a journal collection of his favourite things rather than a thesis. To my mind they are a haphazard collection. Perhaps this reviewer has seen too many cloud-chamber images of sub-atomic particle traces; although we must always remember that those stunningly intriguing patterns — chosen for the cover of the book — are not the particle itself.

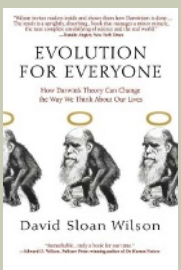
The title's suggestion that this period is the end of representation cannot be further from the truth. At least in science, the end of the twentieth century was the beginning of representation. He is correct in saying: "It strikes me that those makers are most fully preoccupied by whatever cannot be put in a picture," suggesting the representations they are creating are so inadequate that they lead to further questions. Indeed, that is precisely the frustration that has so profoundly informed our present obsession in the laboratory with new and extraordinary visual expressions of science. We are living in an exciting time, in which we continue to ask how we might represent the unrepresentable. ■

Felice Frankel is a senior research fellow in the Faculty of Arts and Sciences at Harvard University's Initiative in Innovative Computing, 60 Oxford Street, Cambridge, Massachusetts 02138, USA. She is co-author of *On the Surface of Things, Images of the Extraordinary in Science*.

Walter Rothschild: The Man, the Museum and the Menagerie
by Miriam Rothschild (Natural History Museum, £9.99)
Walter Rothschild, one of Britain's most famous zoologists and eccentrics, amassed a huge collection of animals. His life, work and family are honestly and affectionately documented by his niece, the late Miriam Rothschild.



Evolution for Everyone: How Darwin's Theory Can Change the Way We Think About Our Lives
by David Sloan Wilson (Delta, $15)
Covering subjects from dung-beetle infanticide to human religion, this book emphasizes the role of group selection in evolution in modern human life. "It is a delicate and subtle debate and Sloan Wilson's popular accounts ... make for enjoyable and thoughtful reading," wrote Mark Pagel (*Nature* **447,** 533; 2007).



Rise of the digital machine

Genomes and language suggest that biological and social complexity emerge from how information is used, argues **Mark Pagel**, not from how much of it there is.

Mark Pagel

Human societies and multicellular organisms share a puzzling feature. They seem to be under-specified. Our societies depend on many more interactions among group members than there are members. Multicellular organisms have many more parts, and connections among those parts, than they have genes. This points to a principle of regulation in the evolution of such complex adaptive systems: complexity arises not from the number of genes or actors but from how those elements are expressed or deployed.

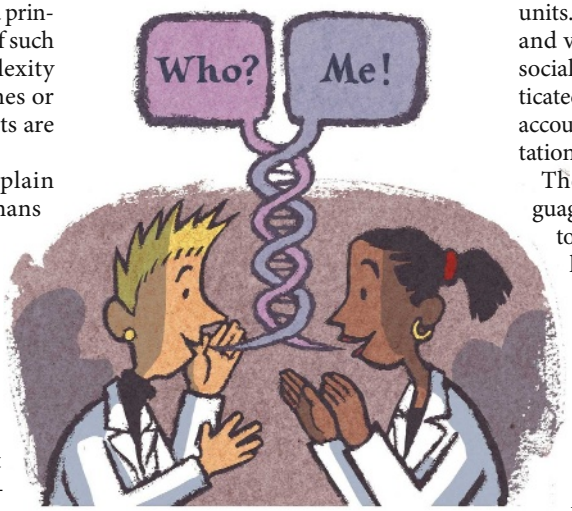
This simple conjecture can explain some remarkable anomalies. Humans are almost unimaginably complex, with trillions of cells organized into hundreds of different tissues. But we have scarcely more genes than a fruitfly or a worm, and only about four or five times as many as brewers' yeast or some bacteria. Surprising then that the human genome is 250 times larger than the yeast's. It comprises about 99% 'junk DNA' — genetic code that is not used to make the protein building blocks of life.

Junk DNA gives every appearance of fulfilling the metaphor of the selfish gene. It accumulates in organisms' genomes simply because it is good at accumulating; it can even be harmful. Why we put up with it has long been a mystery.

Increasingly, it seems that the genes that do code for proteins may recruit some or all of this junk DNA to regulate when, where and how much they are expressed. Because nearly every cell in the body carries a complete copy of the genome, something has to tell the genes that make eyes not to switch on in the back of the head, or genes for teeth to stay silent in our toes. Something has to instruct genes to team up to produce complex structures such as hearts and kidneys, or the chemical networks that create our metabolism and physiology.

Genes, in effect, use regulation to promote their interests within the bodily phenotype: it is how they vary their exposure to the outside world. Regulation is how we can have over 98.5% similarity to chimpanzees in the sequences of our coding genes, yet differ so utterly from them.

Indeed, the huge quantity of junk DNA in the genomes of most complex organisms may act as a vast digital regulatory mechanism. Until recently many common machines, such as aeroplanes, clocks, and even computers were analogue devices, regulated by levers, springs, heat or pressure. Aeroplanes were flown with a stick, springs drove clocks. Digital regulation — instructions encoded in strings of binary numbers arbitrarily long, and hence precise — enabled complexity to increase. Stealth fighters and space shuttles are so complex that they can be flown only by digital computers, not (analogue) human pilots.



D. PARKINS

Similarly, the emergence of digital regulation derived from unused stretches of junk DNA may have precipitated the transition from single cells to complex multicellular organisms. Long runs of the four chemical bases that make up DNA can easily act like binary strings. How these stretches bind to a gene can regulate exquisitely the degree and timing of that gene's expression. Tellingly, bacteria and some other single-celled organisms have negligible amounts of junk DNA. They rely far more on analogue systems of gene regulation that are protein-based and less precise.

A different kind of digital regulation may have made possible the transition to another complex phenotype. Human societies are unique in supporting elaborate systems of cooperation and exchange among unrelated individuals. We help others we may never see again, and we can operate a division of labour — while some hoe the earth, others fetch water. The specialization that cooperation allows is almost certainly responsible for humans' rapid spread out of Africa and around the world. But this kind of social system abounds with ways for cheats to exploit cooperators.

Regulating the social complexity of human societies, then, required something more than the analogue communication systems available to other animals, such as grunts, chest thumping, head butting, odours and bright colours. We need to account for who has done what to whom, when, where and how much they did it, and for what purpose or gain. Human language is a digital communication system of vast possibilities, with words as the discrete units. I suggest it evolved to allow precise and varied regulation of self-interested social behaviour. The exchange of sophisticated verbal information arose to convey accounts of our own and others' acts, reputations, alliances and dues.

The common-sense view is that language arose to promote communication, to coordinate group effort, and to share knowledge. But if the view put forward here is right, language should deliver more of its value from aiding us to manipulate, distort, bewitch and collude than from the cooperative exchange of information. Ants can manage sophisticated societies and deploy large armies, but only we derive substantial benefits from relying on the goodwill of strangers. This depends on how our community judges us and others — verbally.

We use language to promote our interests in society: it is how we vary our expression in the complex social phenotype. As language became more sophisticated in our ancestors it presented more opportunities for selfish gain. We embroider and exaggerate our own dossiers and gently diminish or disparage those of others. We alter our speech or dialect strategically — linguists call it 'code-switching' — to advertise our conformity to particular groups. We can even employ language to enforce moral norms that proscribe or regulate others' social behaviour. Deep down, language may be just the latest form of gene regulation — the voice of our genes. Information management, not lots of parts, is the key to complexity. ■

Mark Pagel is in the School of Biological Sciences, University of Reading, Reading RG6 6AJ, UK, and the Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico, USA 87501.

FURTHER READING

Encode Project Consortium *Nature* **447,** 799–815 (2007).

Gell-Mann, M. *The Quark and the Jaguar: Adventures in the Simple and Complex* (W. H. Freeman, San Francisco, 1994).

Mattick, J. S. *BioEssays* **25,** 930–939 (2003).

Nowak, M. A. & Sigmund, K. *Nature* **437,** 1291–1298 (2005).

NEWS & VIEWS

A. GUENTHER

Pristine rainforest in French Guiana. Global hydrocarbon emissions from tropical rainforests exceed those from cars.

ATMOSPHERIC CHEMISTRY

Are plant emissions green?

Alex Guenther

Hydrocarbon emissions from living vegetation are thought to be harmful to the atmosphere. But the latest study suggests that the negative impact of these emissions in pristine environments is less than expected.

Do trees produce more pollution than cars? Terrestrial vegetation produces copious quantities of hydrocarbons and other volatile organic compounds (VOCs) that dominate global gas emissions. The largest component of emissions from vegetation is the hydrocarbon isoprene. Current models suggest that isoprene emissions in pristine environments can overwhelm the ability of atmospheric oxidants to remove 'greenhouse' gases (such as methane) and toxic gases (such as carbon monoxide). Furthermore, in a polluted atmosphere, isoprene emissions can substantially increase the amount of smog. But on page 737, Lelieveld *et al.*[1] propose that isoprene emitted by terrestrial ecosystems into unpolluted atmospheres is much less deleterious than was previously thought. Only when it comes into contact with air tainted by human activities does it exert a negative influence.

Some VOCs are known to attract pollinators and repel pests, but the function of isoprene remains a mystery, despite its prevalence in biogenic emissions. What was thought to be known was the dramatic negative impact of isoprene on chemical reactions in the troposphere — the lowest part of Earth's atmosphere. The first speculation that abundant hydrocarbon emissions could modify atmospheric composition appeared almost 50 years ago[2]. Modern atmospheric-chemistry model (ACM) simulations now suggest that, in unpolluted air, isoprene depletes the levels of hydroxyl radicals (•OH). Although these radicals are present in the atmosphere in minuscule amounts (less than 1 part per trillion), they are extremely reactive and serve as the primary cleansing agent for the atmosphere. In an atmosphere depleted of •OH, many trace gases would build up to dangerous levels.

Outside urban areas, isoprene is often the dominant reactive-gas emission, and so the typical fate for an •OH molecule in a clean atmosphere is removal by reaction with isoprene. Recent field studies in the remote Amazon basin have provided indirect evidence that isoprene has a considerably smaller effect on •OH than was previously thought[3,4]. Lelieveld *et al.*[1] now confirm this with direct measurements of •OH over the tropical rainforest in Suriname, Guyana and French Guiana. To explain their puzzling observation, the authors propose the existence of a previously overlooked reaction pathway that allows •OH consumed by isoprene to be directly recycled, so minimizing net •OH depletion. By setting the •OH recycling rate to a very high value in a modified ACM, they reproduced observed •OH levels over the rainforest.

Prior to this work[1], there were already indications that something was amiss with the isoprene reactions used in ACMs for pristine atmospheres. For example, the atmospheric isoprene concentrations at several Amazonian sites were found to be lower than expected from the isoprene emissions measured at leaf-level, and from estimates of local •OH concentrations made using a detailed chemistry model[5]. A quick fix could be obtained by lowering the isoprene emission rates used in the model to agree with ambient concentration measurements.

But even these relatively low emission rates were not low enough for global ACMs, which have simplified chemistry models. In those ACMs, such emission rates caused a dramatic depletion of predicted •OH; this in turn leads to unrealistically high predictions of certain key atmospheric constituents. Most global modellers resolve this problem by using isoprene emission rates that are lower than

estimates based on direct emission measurements, and lower than estimates based on the models used for local environments, but that are still within the range of uncertainty[6,7]. A promising approach for developing global maps of isoprene emissions is to make satellite observations of isoprene-oxidation products[8], but this requires an accurate understanding of isoprene-oxidation processes.

The inability of ACMs to incorporate measured isoprene emission rates is not limited to simulations of the undisturbed tropics. Air-quality modellers at US regulatory agencies have also had to decide between using isoprene emission rates based on measurements, or using inaccurate, substantially reduced emission rates that improve the ability of their model to simulate the actual distribution of oxidants[9]. Those modellers opted for the second approach. This might mean that they get the right answers for the wrong reason, but with pressing deadlines for regulatory decisions, they don't have time to wait for scientists to thrash out all the necessary details. Nevertheless, Lelieveld and colleagues' findings[1] should stimulate a re-evaluation of current predictive methods, in which the accuracies of the individual components of an ACM are considered afresh.

None of the recent field campaigns[1,3,4] that looked at isoprene oxidation over remote tropical forests has quantified all of the variables required to adequately constrain model simulations. Lelieveld and colleagues' measurements of •OH concentration are an invaluable step in the right direction, but simultaneous studies of •OH and isoprene fluxes, and of a more comprehensive suite of radicals and isoprene-oxidation products, are also needed. And as the authors point out, laboratory studies of isoprene-oxidation pathways will be essential to understand the processes maintaining •OH levels over tropical forests.

Field studies in less pristine regions are also required to determine whether such oxidation processes are relevant outside Earth's few remaining unpolluted locations. This could lead to a better understanding of the effect of air-quality degradation on the chemical interactions between the biosphere and the atmosphere, and perhaps strengthen arguments for controlling emissions of atmospheric pollutants. ■

Alex Guenther is at the National Center for Atmospheric Research, Earth and Sun Systems Laboratory, 1850 Table Mesa Drive, Boulder, Colorado 80305, USA.
e-mail: guenther@ucar.edu

1. Lelieveld, J. *et al. Nature* **452,** 737–740 (2008).
2. Went, F. W. *Nature* **187,** 641–643 (1960).
3. Karl, T. *et al. J. Geophys. Res.* **112,** D18302 (2007).
4. Kuhn, U. *et al. Atmos. Chem. Phys.* **7,** 2855–2879 (2007).
5. Greenberg, J. P. *et al. Global Change Biol.* **10,** 651–662 (2004).
6. Bey, I. *et al. J. Geophys. Res.* **106,** 23073–23095 (2001).
7. Poisson, N., Kanakidou, M. & Crutzen, P. J. *J. Atmos. Chem.* **36,** 157–230 (2000).
8. Shim, C. *et al. J. Geophys. Res.* **110,** D24301 (2005).
9. Wiedinmyer, C. *et al. J. Geophys. Res.* **110,** D18307 (2005).

IMMUNOLOGY

Blood lines redrawn

Thomas Graf

The generation of blood cells is a complex affair. As the culmination of several years of study by various investigators, the latest research will necessitate revision of textbook accounts of the process.

Red cells, white cells and platelets — on the face of it, the main components of blood look simple enough. All of them are ultimately produced from a common source, haematopoietic stem cells. And the white cells, or leukocytes, have a common general function, that of immune defence. But compared with red cells and platelets, leukocytes come in a large variety of specialized types, produced by a still somewhat mysterious variety of intermediate progenitor cells.

On pages 764 and 768 of this issue, Bell and Bhandoola[1] and Wada *et al.*[2] provide definitive evidence that a central aspect of blood-cell differentiation requires a rethink. Taken together with preceding work, their results show that a previously well-recognized distinction between two developmental lineages — lymphoid and myeloid — does not apply. But to appreciate this news, more details about each of the players and their function are required.

Immune cells are devoted to innate or to adaptive immunity. Components of the innate immune system are natural killer (NK) cells, monocytes/macrophages, granulocytes (a classification that includes neutrophils, eosinophils and basophils), mast cells and dendritic cells. The two arms of adaptive immunity are antibody-producing B cells and two types of T cell, cytotoxic and helper, which are respectively characterized by CD8 and CD4 cell-surface proteins. The well-established thinking has been that, along with NK cells, B and T cells are products of the lymphoid haematopoietic lineage, with all the others — including red cells and platelets — forming from the myeloid lineage. Most blood-cell production occurs in the bone marrow, with the exception of T cells, which originate in the bone marrow but mature in the thymus, a small organ in the upper chest.

The basis for the well-established picture stems from 1997, when a common lymphoid progenitor (CLP) was described[3]; this progenitor could give rise to B and T cells as well as NK cells, but not to the myeloid lineage. A few years later the same investigators identified[4] a common myeloid progenitor (CMP), resulting in a haematopoietic tree that symmetrically branches into lymphoid and myeloid cells; this became the classic scheme of haematopoietic differentiation (Fig. 1). A prediction of this scheme was that CLPs migrate from the bone marrow to the thymus to initiate T-cell development. But this concept was subsequently challenged when it was reported that the predominant thymus-seeding cells do not resemble CLPs but have the characteristics of earlier haematopoietic progenitors[5].

T-cell development in the thymus is astonishingly complex, occurring in up to nine sequential stages[6] that are identified by differences in gene expression and developmental potential. The stages can be simplified into the following succession: early T-cell progenitors (ETP, also called double-negative 1, or DN1, cells); DN2 and DN3 cells; CD4/CD8 double-positive cells; and finally CD4 or CD8 single-positive T cells. The various stages are orchestrated by the different microenvironments to which the cells become exposed as they migrate through the thymus[7].

At which stage do thymic T-cell precursors lose their capacity to differentiate into alternative cell types? In line with the CLP model, mice in which the powerful T-cell regulator Notch1 was inhibited showed an increase in the number of thymic B cells but apparently not of myeloid cells (reviewed in refs 6 and 8). Against the model were several reports showing

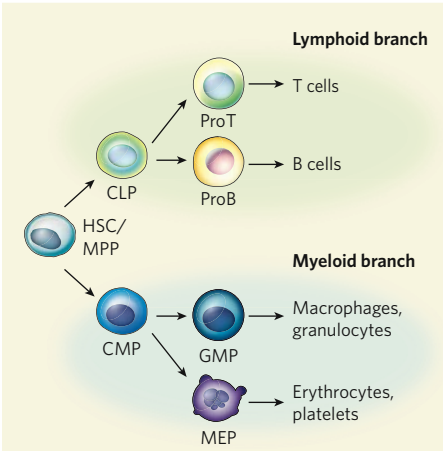


Figure 1 | Classic scheme of haematopoiesis with an early bifurcation into lymphoid and myeloid branches. HSC/MPP, haematopoietic stem cell/multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor. ProT and ProB are progenitor cells that go through several stages to eventually produce T and B cells. GMP, granulocyte macrophage progenitor; MEP, megakaryocyte erythroid progenitor, which gives rise to erythrocytes (red blood cells) and platelets. T cells are specified in the thymus; all other lineages originate in the bone marrow. Arrows indicate cell differentiation.

that the earliest T-cell progenitors have the potential to become macrophages and dendritic cells (reviewed in refs 6, 9 and 10). However, as these experiments were not done clonally under conditions that permit both lymphoid- and myeloid-cell development, the existence of bipotent T-cell/myeloid precursors had not been demonstrated. And even if such progenitors existed, it remained unclear whether they actually participate in the generation of both mature T cells and myeloid cells in the thymus.

Using single-cell assays, Bell and Bhandoola[1] and Wada *et al.*[2] now demonstrate that ETP/DN1 as well as DN2 cells lack B-cell potential, and that a substantial proportion of them have both T-cell and myeloid potential; the latter capacity is lost from the DN3 stage onwards. The myeloid cells in the T-cell/myeloid colonies were found to be predominantly macrophages, but granulocytes[1], dendritic and NK cells[2] were also evident. Intrathymic transplantation experiments[2] of DN1 cells into T-cell-deficient mice showed that about a third of the macrophages were derived from T-cell progenitors; they were detectable in the cortex of the thymus, corresponding to the normal location of thymic macrophages.

In addition, myeloid cells that were purified from the thymus exhibited rearrangements of the genes encoding the business part of the T cell, the T-cell receptor, which are characteristic of ETP/DN1 cells. Finally, a substantial proportion of thymic granulocytes could be lineage-traced[1] to cells that exhibit a particular enzyme activity. The enzyme concerned, RAG recombinase, is central to creating T-cell-receptor diversity, indicating that this myeloid-cell population originates from ETP/DN1 cells in animals.

The new findings leave little room for a physiologically relevant, lymphoid-restricted T-cell progenitor, and call for a revised thymic lineage tree that has a lymphoid–myeloid branching point (Fig. 2). This type of branching seems to be conserved during development, as T-cell/myeloid but not T-cell/B-cell progenitors were found in the fetal liver[2,11]. The existence of lymphoid–myeloid progenitors might also explain the existence of acute myeloid leukaemias that show T-cell-receptor rearrangements. Finally, the revised picture is supported by the remarkable ease with which committed T-cell progenitors can be redirected to become macrophages and dendritic cells using myeloid gene transcription factors[12].

The observation that many T-cell/myeloid progenitors retain dendritic, NK and granulocytic potential raises the question of whether the earliest lineage decisions occur in a sequential fashion or at random. In support of the first possibility is the finding that two subsets of DN2 cells can be distinguished using lineage-tracing experiments, one with and another without the potential to differentiate into dendritic cells[13]. But it is unclear whether subpopulations exist that are restricted in their NK and granulocytic potential.

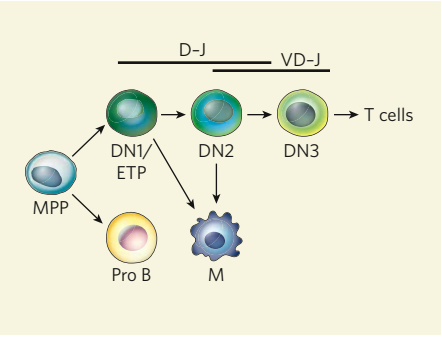


Figure 2 | Revised scheme for thymic T-cell differentiation. The notable difference from Fig. 1 is loss of B-lineage potential before loss of myeloid potential. ETP, early T-cell progenitor, or DN1. DN1 and DN2 are stages of cell differentiation that are subsumed within the ProT category in Fig. 1; DN3 cells belong to a 'PreT' category. The thymus-seeding cells (here designated as MPP) come from the bone marrow through the circulation. Their nature is controversial, and they probably consist of various progenitor subpopulations[5,14,15]. The frequency of transition into macrophages (M) declines from the DN1/ETP stage to the DN2 stage and is lost at the DN3 stage, the T-lineage commitment point. DN1/ETP and DN2 cells have additional differentiation potentials not represented here. The bars underlying D–J and VD–J indicate the rearrangement of T-cell-receptor β-chains and their timing during differentiation.

Developmental immunology has derived great impetus from the simplicity and elegance of a haematopoietic lineage tree symmetrically divided into lymphoid and myeloid branches. But reality is evidently not so neat. The broad lesson to be learned from the revised tree is that it is not always possible to extrapolate from a progenitor's potential to its actual role *in vivo*. A challenge for the future, therefore, is to establish lineage trees that faithfully represent the predominant developmental flows *in vivo*, both in qualitative and quantitative terms. Getting lineage relationships right is essential for successfully modelling the changes in the networks of transcription factors and molecular landscapes that occur as precursors differentiate into mature cells. Bell and Bhandoola[1] and Wada *et al.*[2] have brought T-cell development a step closer to becoming one of the premier systems in which this goal may be achieved. ■

Thomas Graf is at the Center for Genomic Regulation, Carrer Dr Aiguader 88, 08003 Barcelona, Spain.
e-mail: thomas.graf@crg.es

1. Bell, J. J. & Bhandoola, A. *Nature* **452,** 764–767 (2008).
2. Wada, H. *et al. Nature* **452,** 768–772 (2008).
3. Kondo, M., Weissman, I. L. & Akashi, K. *Cell* **91,** 661–672 (1997).
4. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. *Nature* **404,** 193–197 (2000).
5. Bhandoola, A., von Boehmer, H., Petrie, H. T. & Zúñiga-Pflücker, J. C. *Immunity* **26,** 678–689 (2007).
6. Rothenberg, E. V., Moore, J. E. & Yui, M. A. *Nature Rev. Immunol.* **8,** 9–21 (2008).
7. Petrie, H. T. & Zúñiga-Pflücker, J. C. *Annu. Rev. Immunol.* **25,** 649–679 (2007).
8. Maillard, I., Fang, T. & Pear, W. S. *Annu. Rev. Immunol.* **23,** 945–974 (2005).
9. Ceredig, R. & Rolink, T. *Nature Rev. Immunol.* **2,** 888–897 (2002).
10. Laiosa, C. V., Stadtfeld, M. & Graf, T. *Annu. Rev. Immunol.* **24,** 705–738 (2006).
11. Katsura, Y. *Nature Rev. Immunol.* **2,** 127–132 (2002).
12. Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. *Immunity* **25,** 731–744 (2006).
13. Masuda, K. *et al. J. Immunol.* **179,** 3699–3706 (2007).
14. Iwasaki, H. & Akashi, K. *Immunity* **26,** 726–740 (2007).
15. Lai, A. Y. & Kondo, M. *Proc. Natl Acad. Sci. USA* **104,** 6311–6316 (2007).

MATERIALS SCIENCE

Strong teeth, strong seeds

Peter Ungar

A full account of the relationships between tooth form, structure and function remains out of reach. Viewing teeth from an engineering materials perspective offers a way to help crack the problem.

Thickened tooth enamel is a trait that separates most early hominins (our distant ancestors and cousins) from the chimpanzees and gorillas (our nearest living relatives). According to conventional wisdom, our forebears came out of the trees as the savannahs spread through eastern and southern Africa between about 2.5 million and 1.5 million years ago. The story goes that this change of venue was accompanied by a change in diet from soft, fleshy forest fruits to hard, brittle roots, tubers or other savannah foods. An open setting also meant more grit and an increasingly abrasive diet. Thus, enamel that was thickened to resist breakage and wear is seen as a milestone on the road to humanity. Although relationships between tooth-crown height and habitat type have been proposed for some mammals[1], and associations between enamel thickness and food hardness have been documented for others[2], there is no overarching body of theory to explain the evolution of this trait.

Writing in *BioEssays*, Lucas *et al.*[3] address this problem in mammalian evolution from the perspective of engineering materials science. They use the principles of fracture and deformation mechanics to predict the thickness and distribution of enamel on the crowns of

NATURE|Vol 452|10 April 2008

50 & 100 YEARS AGO

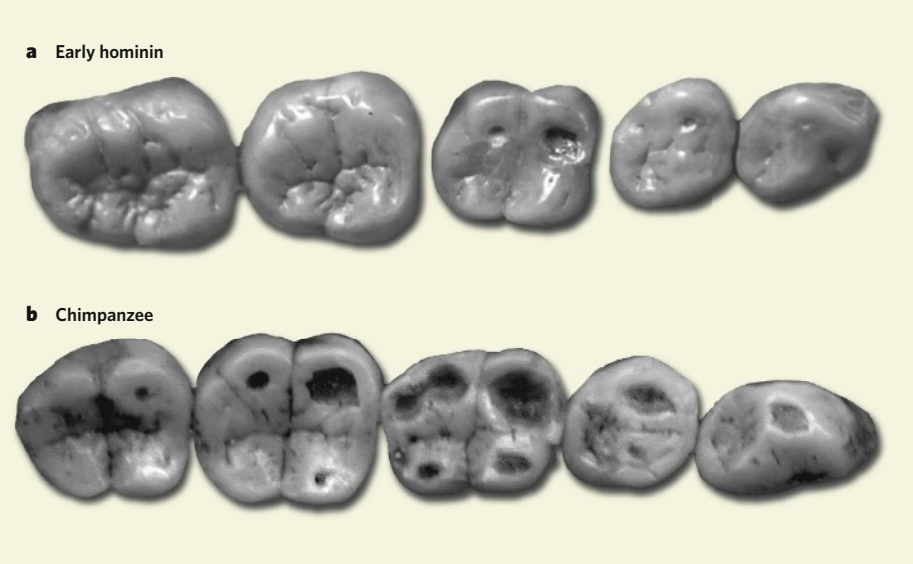
50 YEARS AGO

"A Suggested Revision of Nomenclature — Angiotensin" The vasoactive peptide resulting from the action of rennin on an alpha-globulin was discovered by two groups of investigators, with the result that it received two trivial names, angiotonin and hypertensin. Synthesis of the octapeptide has now confirmed the identity of this peptide and justifies dropping the double nomenclature. We propose the simplified name, angiotensin, and its derivatives angiotensinase and angiotensinogen. Angiotensin is a hybrid word but does, we think, have the advantage of being easy to pronounce even with a variety of accents, is euphonious, and is understandable despite the most recalcitrant microphone. **Eduardo Braun-Menendez & Irvine H. Page**
From *Nature* 12 April 1958.

100 YEARS AGO

Paris, Academy of Sciences, March 30 ... The determination of time, both on land and at sea, with the aid of wireless telegraphy: Bouquet de la Grye. With the present installation at the Eiffel Tower, wireless signals can be sent a distance of 2000 kilometres, and it has been estimated that by increasing the electric energy this distance could be doubled. It is suggested that a special signal should be sent exactly at midnight ... Such a time signal would be of the greatest service to navigators within its radius. On the proposal of the president, the examination of the proposition ... was referred to a committee composed of the members of the sections of astronomy, geography, navigation, and physics, together with MM. Darboux, Poincaré, and Cailletet.

A phenomenon attributable to positive electrons in the spark spectrum of yttrium: Jean Becquerel. The experimental study of the Zeeman effect in the spark spectrum of yttrium gives results which may be most simply explained by assuming the presence of positive electrons.
From *Nature* 9 April 1908.



P. UNGAR

Figure 1 | Shear quality. Comparisons of moderately worn premolar and molar crown surfaces of (**a**) an early hominin, *Australopithecus afarensis* (fossil AL266, dating to about 3.3 million years ago), and (**b**) a chimpanzee. Wear of thin enamel, as seen in the chimpanzee, quickly leads to sharp edges where the hard outer surface gives way to softer, underlying dentine. The hominin molar is flatter, with thicker enamel, and wear generally does not produce sharp edges. Chimpanzee teeth may be better suited to efficient shearing or grinding of softer, tougher foods than to crushing of hard, brittle items.

premolars and molars, given food with specific fracture properties. The principal job of a tooth is to fracture foods, and teeth can be stressed by considerable, repeated loading forces as they go about the business of chewing. Teeth must break foods without themselves being broken; they do not 'heal', and the teeth of most adult mammals cannot be replaced.

Starting from these premises, Lucas *et al.* offer several predictions. The first relates to the fact that tooth crowns are bilayered, with hard, brittle enamel overlying more compliant dentine. A thin enamel coat would be more prone to flex and cause tensile stresses within the tooth, leading to cracks that would start at the undersurface of the crown and spread outwards. Thicker enamel would provide one way to resist tooth fracture caused by the high stresses associated with feeding on hard objects. Changing the structure of the enamel would also help. Enamel forms as rods or prisms that align together like bunches of dry spaghetti strands. The cap is better able to resist forces oriented parallel to the long axis of these prisms than perpendicular to them. Furthermore, cracks spreading between adjacent prisms may be stopped if the rods change direction periodically (a phenomenon known as decussation).

Lucas *et al.* propose decussation as a mechanism for strengthening teeth that have enamel of uneven thickness. As an example of another strengthening mechanism, soft foods would be expected to 'smother' a tooth surface, redistributing tensile forces to the margins. This might explain why some mammals possess a cingulum, a collar of enamel around the periphery of the crown.

The authors also consider fracture resistance from the food's perspective. It does a plant little good to have its seeds destroyed, so seed coats should also evince adaptations to resist breakage. Such adaptations can be remarkably similar to those seen in teeth, given that both tooth crowns and seed shells are often bilayered. This leads to predictions for shell thickness and internal-fibre decussation that are analogous to those seen in teeth. Tooth–food interactions can thus be viewed as a 'death match' as nature selects for strength in both dental enamel and seed shells.

The application of principles from fracture and deformation mechanics to address the relationships between tooth and food structures opens avenues for productive research. There have been debates about the best way to measure enamel thickness, given its uneven distribution across a crown[4,5]. A materials-science approach, combined with new imaging technologies such as micro-computed tomography, could help in the development of a standard, functionally relevant protocol. It could also help in formulating computer simulations known as finite-element models that attempt to illuminate the relationships between tooth form and chewing stress[6], as both methods aim to elucidate how teeth and foods react when the jaws close on an item of food.

Teeth are highly complicated structures, and much remains to be done if we are to understand the adaptive significance of mammalian dental form. In terms of enamel structure, functionally relevant aspects of complexity occur at many different levels: crystallites form prisms, prisms are organized into enamel types, and enamel types are combined into schmelzmusters, or layers[7]. These must ultimately be built into the models

relating tooth anatomy to function.

Further, enamel thickness (or more precisely, thinness) relates to more than just resistance to fracture. The spatial relationship between the harder enamel cap and the softer 'horns' of dentine, which underlie the tooth cusps, results in sculpting of the tooth surface with wear. Enamel can be thinner in certain areas to facilitate wear, generating sharp edges for efficient shearing where the two tissue types meet[8] (Fig. 1). The crescent-shaped crowns of camels and ruminants provide an example of how the distribution of enamel and dentine can combine with wear to produce efficient grinding surfaces. And it's not only enamel thickness that contributes to crown sculpting: aspects of enamel microstructure, such as prism alignment, can affect rates of wear[9], as can dentinal structure[10].

We have a great deal more to do in explaining the diverse array of factors affecting the relationships between tooth form, structure and function. Nevertheless, the potential for understanding the palaeobiology of early hominins and other fossil mammals certainly justifies the efforts. The problems posed by dental functional anatomy are complex, but Lucas *et al.*[3] have produced an elegant and exciting new approach for tackling them. ■

Peter Ungar is in the Department of Anthropology, University of Arkansas, Fayetteville, Arkansas 72701, USA.
e-mail: pungar@uark.edu

1. Macho, G. A. & Spears, I. R. *Am. J. Phys. Anthropol.* **109,** 211–227 (1999).
2. Dumont, E. R. *J. Mammal.* **76,** 1127–1136 (1995).
3. Lucas, P., Constantino, P., Wood, B. & Lawn, B. *BioEssays* **30,** 374–385 (2008).
4. Shimizu, D., Macho, G. A. & Spears, I. R. *Am. J. Phys. Anthropol.* **126,** 427–434 (2005).
5. Teaford, M. F. in *Evolution of the Human Diet: The Known, the Unknown and the Unknowable* (ed. Ungar, P. S.) 56–76 (Oxford Univ. Press, 2007).
6. Kono, R. T. *Anthropol. Sci.* **112,** 121–146 (2004).
7. von Koenigswald, W. & Clemens, W. A. *Scanning Microsc.* **6,** 195–218 (1992).
8. Ungar, P. S. & M'Kirera, F. *Proc. Natl Acad. Sci. USA* **100,** 3874–3877 (2003).
9. Rensberger, J. M. & von Koenigswald, W. *Paleobiology* **6,** 477–495 (1980).
10. Kierdorf, H. & Kierdorf, U. *Anat. Embryol.* **186,** 319–326 (1992).

QUANTUM PHYSICS

Observations turn up the heat

Kimberly R. Chapin and Marlan O. Scully

The idea that observers can influence what they observe has a history that stretches back beyond quantum physics. That we can affect how a system heats up and cools down simply by probing it is a new twist.

As the great quantum physicist Werner Heisenberg — he of the uncertainty principle — made plain, in quantum mechanics, separation of the observer from the phenomenon to be observed is not possible. But in fact, the strange idea that consciousness, intelligence and the act of observation are intertwined with physical phenomena predates Heisenberg. Specifically, James Clerk Maxwell famously introduced into his studies of thermodynamics "a being whose faculties are so sharpened that he can follow every molecule in its course"[1], such that it could identify and siphon off the hotter (faster) molecules in a gas[2]. 'Maxwell's demon' would thus be able to extract useful work from the system, while heat is in effect transferred from a cooler to a hotter region — in clear breach of the normal direction of heat flow from hotter to cooler encapsulated in the second law of thermodynamics.

On page 724 of this issue, Erez *et al.*[3] provide a neat link between these physical curiosities, by suggesting a way to use the quantum measurement process to control a system's thermodynamics, in the spirit of Maxwell's demon. At the heart of their concept is the quantum-physical equivalent of the old adage 'a watched pot never boils'. This is the quantum Zeno effect, which states that, if you measure a quantum system often enough, it will never be able to change its state, and so will not evolve at all.

The authors consider disturbances to the thermal equilibrium of a system of two energy levels surrounded by an infinite reservoir of heat. They show that the entropy and temperature of both the system and its surrounding 'heat bath' can be increased or decreased by changing how frequently the system is probed. In principle, therefore, it should be possible to control rapidly both the production of entropy and the heating or cooling of the system just by changing how we probe it. It might even be feasible to make heat flow from a colder region to a hotter region for short times.

These considerations[3] are a nice illustration of how the historical development of physics has been turned on its head. Quantum mechanics first saw the light of day when Max Planck began to study the thermodynamics of heat absorbed and emitted by a perfect 'blackbody' radiator. But now it is our continuously developing understanding of quantum physics that is extending and improving our conception of thermodynamics. This is more than an esoteric consideration. The interaction has brought many practical benefits — not least, the development and improvement over the past 50 years of the lasers that are at the heart of so much modern technology.

Perhaps the first instance of this fruitful interaction is a seminal 1956 paper by Norman Ramsey[4]. In it, he showed how negative thermodynamic temperatures are possible in quantum-mechanical systems consisting of two energy levels. The negative-temperature state corresponds to 'population inversion': the upper energy state is more densely populated than the lower, the reverse of the norm. Ramsey went on to observe that, in the common Kelvin–Planck formulation of the second law of thermodynamics, any amount of work can be converted to heat, but heat cannot be converted completely to useful work. Yet in a

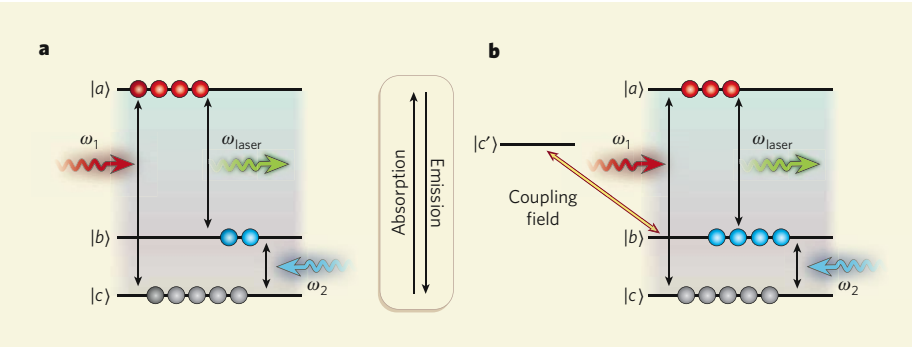


Figure 1 | Laser quantum heat engines. a, With population inversion; **b**, without. **a**, In a conventional laser, atoms from the ground state $|c\rangle$ of the lasing medium are pumped into a higher-energy state, $|a\rangle$, by energetic, incoherent pump radiation of frequency ω_1. Meanwhile, an intermediate energy state $|b\rangle$ is coupled to the ground state by less-energetic light of frequency ω_2. If more atoms can be pumped into $|a\rangle$ than into $|b\rangle$, a population inversion is created, with more atoms in the higher-energy state. This is the precondition for the emission of coherent laser light as atoms in state $|a\rangle$ de-excite. A high degree of population inversion is required, as some of the light emitted from $|c\rangle$ will be immediately reabsorbed, re-exciting the system. **b**, In a laser without population inversion, an external laser field couples a further energy state, $|c'\rangle$, to the system. This coupling causes a quantum interference effect that serves to suppress the reabsorption of light, while leaving emission unaffected. As Erez *et al.* show[3], a similar but subtler control of the thermodynamics of an energy-level system is possible by varying the rate of measurement (the frequency with which a coupling field is applied) and exploiting the quantum Zeno effect.

population-inverted system, the opposite is true: any amount of heat can be turned into work, but work cannot be completely converted into heat.

The first lasers worked along these lines, producing work (coherent laser light) from heat (an inverted atomic population in the lasing medium; Fig. 1a). A feature of this type of laser is that energy, in the form of incoherent light, must be put into the lasing medium to establish a population inversion. Atoms in the ground state absorb that light; atoms in the excited state amplify it. Lasing will not start, therefore, until we have pumped a 'threshold' energy into the system to propel the majority of atoms into an excited state. As light emitted from excited-state atoms can also be reabsorbed by ground-state atoms, there has to be a considerable degree of inversion for the laser to operate.

But it turns out that this 'threshold problem' need not be a problem: again, it was quantum-mechanical considerations that pointed the way to the realization that laser emission could be possible even without population inversion[5,6]. The precondition is that absorption can be suppressed and emitted photons are not reabsorbed, so that every excited-state atom can contribute to the emission of laser light. This has been achieved experimentally[7] using a coupling laser field as a kind of continuous probe, the effect of which is to suppress absorption (Fig. 1b).

Erez and colleagues' measurement probe is nothing less than a rapid sequence of ultrashort pulses. Their suggestion[3] that we might be able to control the thermodynamics of an atomic system simply by altering the way in which it is measured connects directly with the principle of lasing without inversion. With such subtle and rapid control, better manipulation of absorption and gain in lasers might be at hand. One pay-off might be lasers operating at X-ray frequencies, at which, owing to the rapid decay of excited states, it is currently difficult to prepare a population inversion. ■

Kimberly R. Chapin and Marlan O. Scully are at the Institute for Quantum Studies, Texas A&M University, College Station, Texas 77843, USA. Marlan O. Scully is also in the Applied Physics and Materials Science Group, Princeton University, Princeton, New Jersey 08544, USA.
e-mails: k-chapin@tamu.edu; scully@tamu.edu

1. Maxwell, J. C. *Theory of Heat* (Longman's Green, London, 1871).
2. Scully, R. & Scully, M. *The Demon and the Quantum: From the Pythagorean Mystics to Maxwell's Demon and Quantum Mystery* (Wiley-VCH, Weinheim, 2007).
3. Erez, N., Gordon, G., Nest, M. & Kurizki, G. *Nature* **452,** 724–727 (2008).
4. Ramsey, N. *Phys. Rev.* **103,** 20–28 (1956); see also Ch. 7 of ref. 2.
5. Kocharovskaya, O. *Phys. Rep.* **219,** 175–190 (1992).
6. Harris, S. E. *Phys. Today* **50,** 36 (1997).
7. Zibrov, A. *et al. Phys. Rev. Lett.* **75,** 1499–1502 (1995).

CELL BIOLOGY

Porter and sorter

Michael G. Roth

Clathrin is a protein familiar for its ability to import material into cells. But it also seems to mediate another crucial process — helping newly made proteins to pick the right destination on the cell surface.

The plasma membrane of differentiated cells, such as sperm, neurons and epithelial cells, is divided into specialized surfaces with distinct protein and lipid compositions. For example, the plasma membrane of epithelial cells is separated by impermeable tight junctions into so-called apical and basolateral domains, which differ considerably in their structure, composition and function. The apical membrane faces the lumen of an organ and contains transporters for importing materials into, or exporting them out of, the body. The basolateral membrane faces other cells forming the organ, and contains transporters for the uptake of nutrients, as well as adhesion proteins for attaching to neighbouring cells.

It is known that specific mechanisms ensure the sorting of newly synthesized proteins to, and their retention at, their correct membrane destination[1]. But the precise mechanisms responsible are not known. On page 719 of this issue, Deborde and colleagues[2] now show that clathrin, which plays an essential part in endocytosis (a process through which molecules and other materials enter the cell), also has a direct function in sorting proteins destined for the basolateral membrane.

Early insight into the mechanisms responsible for the non-random distribution of membrane proteins came from a study of the maturation of enveloped viruses in epithelial cells[3]. These viruses become enclosed in an outer coat made from the budding-off of a small piece of their host cell's plasma membrane. In epithelial cells, certain viruses assemble at the apical membrane and others at the basolateral membrane, but the same type of virus never buds from both membranes.

Extensive work that followed[4–6] showed that proteins destined for the plasma membrane, including viral envelope proteins, are somehow identified by short sequences of only a few amino acids called basolateral sorting signals. Curiously enough, some of the same signals also serve as recognition sequences for clathrin-mediated endocytosis once the proteins arrive at the cell surface. This finding implied that clathrin might be involved in protein sorting at the trans-Golgi network (TGN) — the part of a membrane-bound molecular-packaging organelle known as the Golgi complex at which, on the way to their final destination, apical and basolateral glycoproteins part company[1] (Fig. 1). Remarkably, in the 17 years since this discovery, no direct evidence for clathrin's role in the sorting of basolateral proteins at the TGN has been obtained until now[2].

The lack of direct evidence has been frustrating. After all, the assembly of clathrin-coated membrane vesicles during endocytosis is the best-understood mechanism for the formation of cellular transport vesicles. At the cell surface, transmembrane cargo molecules, which bear short peptide motifs in their cytoplasmic portions, are recognized by adaptor proteins in the cell membrane. Clathrin molecules also bind to these adaptors and polymerize into a curved array that deforms the membrane into a spherical vesicle (Fig. 1). With the help of accessory proteins, the clathrin coat pinches off the membrane, carrying the cargo package to its next cellular destination[7].

Does clathrin participate in the sorting of basolateral proteins in the same way? One line of thought has been that, as there is a set of physical requirements for collecting the correct cargo and deforming membranes into an intracellular transport vesicle, the vesicles must all form in the same way. But with no direct evidence for clathrin's role in the formation of vesicles that transport basolateral proteins, researchers were left wondering whether there is any similarity between this process at the TGN and the formation of clathrin-coated vesicles at the plasma membrane during endocytosis.

Deborde *et al.*[2] set out to answer this question. Using a technique known as RNA interference (RNAi) in the MDCK epithelial cell line, they reduced clathrin levels by more than 90%. This treatment had the expected effect: it slowed the best-known non-endocytic process mediated by clathrin — the exit of lysosomal enzymes from the TGN. Moreover, although tight junctions were unaffected, the reduction in clathrin levels led to a random distribution of basolateral, but not apical, proteins.

Next, the authors designed genes in which either basolateral or apical proteins were fused to fluorescent proteins. They expressed these genes in clathrin-depleted cells under conditions that allowed them to visualize a specific, newly synthesized protein in real time during its transport from the TGN. At low clathrin levels, the exit rate of the basolateral fluorescent proteins from the TGN decreased to that of the apical 'reporter' proteins, and 53% of the basolateral fluorescent proteins were sorted into the transport vesicles containing the apical fluorescent reporter. These observations indicated that, although clathrin plays a significant part in the sorting of basolateral proteins, in its absence a direct but inefficient

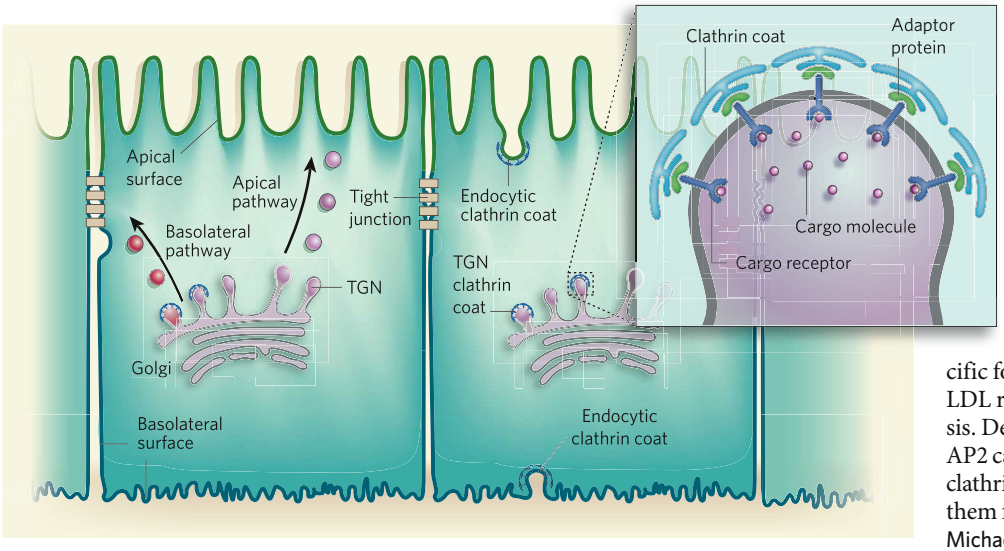


Figure 1 | Clathrin, a multifaceted protein. Epithelial cells, which are attached to their neighbours by impermeable tight junctions, contain specialized apical and basolateral membrane surfaces; tight junctions prevent diffusion of membrane lipids and proteins between these surfaces. As in many other cell types, clathrin mediates endocytosis in epithelial cells. Moreover, on the trans-Golgi network (TGN), clathrin coats are known to be essential for protein transport to other cellular compartments involved in intracellular trafficking[8]. Deborde *et al.*[2] now show that clathrin is also necessary for the sorting of basolateral proteins at the TGN.

transport pathway from the TGN to the basolateral surface might continue to operate. But is this really the case?

A noteworthy limitation of RNAi is that it often takes so long for a protein to degrade once its synthesis has been blocked by the RNAi machinery that it becomes difficult to distinguish the direct effects of that protein's depletion from its indirect effects. For example, loss of clathrin would stop clathrin-mediated endocytosis, and this might prevent a protein necessary for basolateral vesicles from returning from the basolateral surface to the TGN. So Deborde *et al.* used a different method to rapidly inactivate clathrin. They expressed a form of clathrin in MDCK cells that aggregates (and thus becomes inactive) in the presence of a crosslinker molecule. With clathrin aggregation, the exit from the TGN of two different basolateral proteins, carrying different basolateral sorting signals, was delayed, whereas the transport rate of an apical protein was unaffected. This completely different approach supports the simplest interpretation of the RNAi experiments, that clathrin is directly necessary for basolateral sorting at the TGN.

Having established a role for clathrin in the sorting of basolateral proteins, Deborde and colleagues looked at whether this process is linked to clathrin-mediated endocytosis. Depletion of another essential component of the endocytic pathway, the clathrin adaptor protein AP2, severely inhibited endocytosis, but had no effect on the kinetics of basolateral-protein export from the TGN. So clathrin's function in basolateral-protein sorting seems to be independent of its role in endocytosis.

But clathrin coats have been observed on TGN membranes, where they are essential for other aspects of intracellular trafficking[8]. So it is likely that, for the sorting of basolateral proteins at the TGN, clathrin coats collect cargo in much the same way as they do during endocytosis at the plasma membrane. If so, this indicates that there is much to be discovered about basolateral sorting. Many different cargo-specific adaptor proteins mediate the clathrin–cargo interaction at the plasma membrane, thereby allowing the entry of different classes of cargo irrespective of vast differences in their concentrations. Adaptor proteins that would recognize different classes of basolateral sorting signals at the TGN have not been identified, but they are likely to fall into classes, such as adaptors specific for G-protein-coupled receptors and the LDL receptor family, as is seen for endocytosis. Deborde and colleagues' experiment with AP2 can be applied systematically to all known clathrin adaptor proteins to determine if any of them function at the TGN. ■

Michael G. Roth is in the Department of Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9038, USA.
e-mail: michael.roth@utsouthwestern.edu

1. Rodriguez-Boulan, E., Kreitzer, G. & Musch, A. *Nature Rev. Mol. Cell Biol.* **6,** 233–247 (2005).
2. Deborde, S. *et al. Nature* **452,** 719–723 (2008).
3. Rodriguez-Boulan, E. & Sabatini, D. D. *Proc. Natl Acad. Sci. USA* **75,** 5071–5075 (1978).
4. Brewer, C. B. & Roth, M. G. *J. Cell Biol.* **114,** 413–421 (1991).
5. Matter, K., Hunziker, W. & Mellman, I. *Cell* **71,** 741–753 (1992).
6. Hunziker, W. & Fumey, C. *EMBO J.* **13,** 2963–2969 (1994).
7. Roth, M. G. *Nature Rev. Mol. Cell Biol.* **7,** 63–68 (2006).
8. Bard, F. & Malhotra, V. *Annu. Rev. Cell Dev. Biol.* **22,** 439–455 (2006).

NEURODEGENERATION

A question of balance

Leslie Michels Thompson

When a disease-associated gene is mutated, is the cellular activity of its protein product enhanced or reduced? For at least one neurodegenerative disease, spinocerebellar ataxia 1, the answer seems to be both.

Ten devastating neurodegenerative disorders, including Huntington's disease, that are caused by mutations in unrelated genes have one feature in common: in the disease gene, the CAG nucleotide triplet, which encodes the glutamine amino acid, is repeated many times. Consequently, the resulting protein contains a tract of glutamine residues, and disorders with this feature are known collectively as polyglutamine-repeat diseases[1]. But although the polyglutamine chain is toxic to cells, it is not in itself sufficient to cause human disease[1–3]. So what are the other molecular players, and how do they interact to bring about neurodegeneration? To address these questions, Lim *et al.*[4] (page 713 of this issue) extended the search for molecular partners of ataxin 1 (ATXN1), which when mutated is responsible for the polyglutamine disease spinocerebellar ataxia 1, and tested whether interactions of ATXN1 with these partners is altered in disease. They find that ATXN1 mutation is accompanied by alterations in the cellular levels of at least two distinct ATXN1-containing, disease-associated protein complexes.

There are two main types of mutation. In the less common gain-of-function mutations, the protein concerned either acquires new activity or its existing activity is enhanced. By contrast, in loss-of-function mutations, the activity of the mutated protein is either reduced or abolished. In polyglutamine diseases, because the extra glutamine residues in the mutant protein acquire toxic properties — through impaired protein folding and degradation pathways, altered subcellular localization, abnormal interactions with other cellular proteins and other events — early research focused on target proteins and cellular processes that contribute to a gain of function. The possibility that loss-of-function effects play a part in these diseases was considered less likely: knock-out mice that

NATURE|Vol 452|10 April 2008

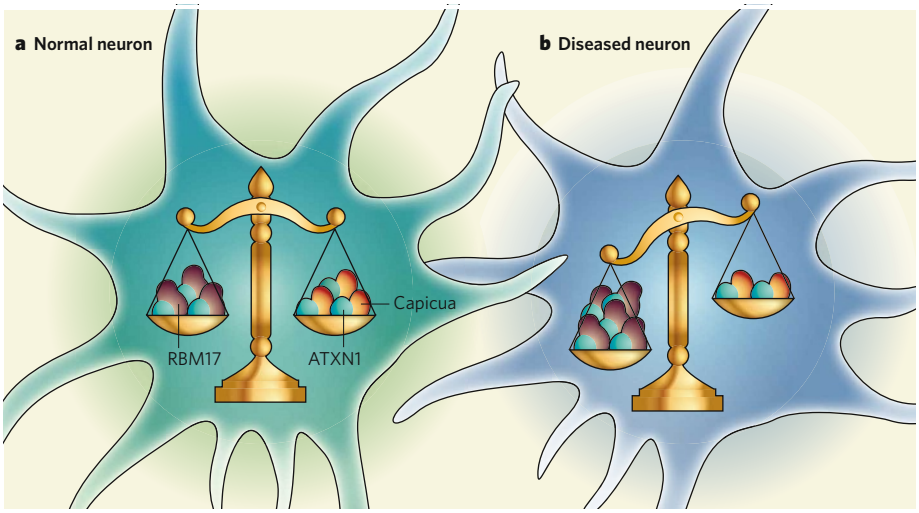


Figure 1 | Protein interactions implicated in spinocerebellar ataxia 1. Lim *et al.*[4] find that an expanded glutamine repeat within the mutated ATXN1 protein shifts the balance of cellular protein complexes. The levels of complexes that ATXN1 makes with each of two other cellular proteins, the neurotoxic RBM17 and the neuroprotective capicua, seem to vary depending on whether a neuron expresses normal ATXN1 (**a**) or its mutated version (**b**). So when ATXN1 is mutated, a simultaneous increase in the levels of RBM17–ATXN1 complexes and decrease in those of capicua–ATXN1 complexes disrupt the critical balance essential for proper neuronal function.

do not express a polyglutamine-disease gene, including *Atxn1*, do not have the disease symptoms that are observed when these genes are mutated, and nor do humans with deletions of such genes[5,6]. But, although it is tempting to conclude from these previous studies that loss of function does not have significant consequences, a more likely explanation is that, whereas the normal function of a disease gene is necessary, its loss is not sufficient to cause a polyglutamine disease.

With the exception of spinal and bulbar muscular atrophy[2], the culprit protein for most polyglutamine disorders was first characterized as of unknown function. For example, it emerged only some time after its discovery that the huntingtin protein, which is mutated in Huntington's disease, is likely to regulate axonal trafficking in neurons, enhance neuron-specific gene transcription by sequestering the transcription-repressor protein REST, and potentially act as an essential scaffolding protein to organize the components of cellular pathways or structures[3,7]. With polyglutamine-repeat expansion in the mutant huntingtin, and when this is mimicked experimentally[8], REST has greater freedom to suppress gene expression; mutant huntingtin might even aid this process. Similar loss-of-function effects are observed after mutation of the ataxin 7 or ataxin 3 genes, which lead to other polyglutamine diseases[1,2].

For spinocerebellar ataxia 1, several studies[1,6] implicated both gain-of-function of the mutant ATXN1 and loss-of-function of the normal protein in its pathology. Previous work[1,6] also established a disease-associated role for specific domains within ATXN1 other than the polyglutamine repeat; in different polyglutamine diseases, modifications of different protein domains seem to be involved in modulation of disease.

Lim *et al.*[4] show that the polyglutamine repeat alters the propensity of specific protein complexes to form in mouse neurons, and that components of these complexes compete for interaction with ATXN1. They find that the length of a repeated glutamine tract in ATXN1, which in disease expands to 39 repeats or more, influences the distribution of cellular proteins in ATXN1 complexes. When the tract expands, the formation of complexes containing a protein that promotes neurotoxicity, RBM17, is enhanced. The authors also show that, in the mouse brain, RBM17 is highly expressed in Purkinje cells of the cerebellum — the region most affected in spinocerebellar ataxia 1 — and that in a fruitfly model of spinocerebellar ataxia 1, a reduction in RBM17 levels protects these insects from retinal degeneration; increased levels of this protein promote degeneration. So the increased formation of ATXN1–RBM17 complexes may represent a gain of function that seems to contribute to the neurodegeneration observed in spinocerebellar ataxia 1.

Lim and colleagues further find that when ATXN1 is mutated, the formation of complexes between it and a protein called capicua is reduced. Previously, the authors have shown[6] that high levels of capicua suppress neurotoxicity in flies. So with decreased formation of ATXN1–capicua complexes in the brain of animals with spinocerebellar ataxia 1, the protective effect of capicua might be lost. Mice that do not express normal ATXN1 but carry a single copy of the mutant gene have worse motor skills and shorter lifespans than mice that express the normal gene as well as a copy of the mutant one. These observations support the authors' proposed mechanism that normal ATXN1 promotes the formation of a protective complex and keeps the formation of a harmful one in check.

Lim and colleagues' data do not directly prove that the redistribution of ATXN1 between RBM17-containing and capicua-containing complexes causes spinocerebellar ataxia 1. But together, these observations indicate that, on ATXN1 mutation, a reduction in native complexes containing normal ATXN1 occurs simultaneously with a rise in a harmful complex containing mutated ATXN1 (Fig. 1). Such protein-interaction studies are likely to lead to insights relating to disease mechanisms, although the answers won't always be straightforward.

The complexity of such disorders is illustrated by a protein-interaction screen that identified more than 200 protein partners for mutant huntingtin[9]. Of these, 60 genes encoding the interactor were tested for their ability to behave as genetic modifiers of neurodegeneration in a fruitfly model, and as many as 45% of these had a modulatory effect on the phenotype of Huntington's disease. It will be intriguing to determine whether these interacting proteins are expressed in the affected regions of the Huntington-diseased brain and to compare the strength of these interactions and the distinct complexes that may form in tissues expressing the mutant protein with those expressing normal huntingtin, as shown here for ATXN1.

Proving a direct link between protein interactions and disease will be the next challenging step. One approach might be to perform experiments similar to those of Lim *et al.* for complexes that form preferentially in the presence of mutant polyglutamine-repeat proteins. But causality will require the expression of relevant disease proteins with mutations in specific domains that prevent complex formation, or other mutations that prevent these complexes from forming *in vivo*. Unlike neurodegenerative diseases with complex genetics, single-gene disorders such as polyglutamine diseases allow researchers to address specific questions relating to the impact of the polyglutamine-repeat expansion on the function of the normal and mutant proteins. Ultimately, such studies should tease out molecules crucial for the maintenance of protein balance in the brain. ■

Leslie Michels Thompson is in the Departments of Psychiatry and Human Behavior, of Neurobiology and Behavior and of Biological Chemistry, University of California, Irvine, Gillespie 2121, Irvine, California 92696-4260, USA.
e-mail: lmthomps@uci.edu

1. Orr, H. T. & Zoghbi, H. Y. *Annu. Rev. Neurosci.* **30,** 575–621 (2007).
2. Katsuno, M. *et al. Exp. Neurol.* **200,** 8–18 (2006).
3. Li, S. & Li, X.-J. *Mol. Neurodegen.* **1,** 19 (2006).
4. Lim, J. *et al. Nature* **452,** 713–718 (2008).
5. Bates, G. P., Mangiarini, L. & Davies, S. W. *Brain Pathol.* **8,** 699–714 (1998).
6. Lam, Y. C. *et al. Cell* **127,** 1335–1347 (2006).
7. Shao, J. & Diamond, M. I. *Human Mol. Genet.* **16,** R115–R123 (2007).
8. Zuccato, C. *et al. J. Neurosci.* **27,** 6972–6983 (2007).
9. Kaltenbach, L. S. *et al. PLoS Genet.* **3,** e82 (2007).

Q&A

CELL BIOLOGY

SUMO

Erik Meulmeester and Frauke Melchior

A protein called small ubiquitin-related modifier (SUMO) can be coupled to other proteins to control their function. This SUMOylation has been implicated in the regulation of a host of cellular processes, and is essential for the health, and even the survival, of most organisms.

Is SUMO unique in the way it regulates proteins?

No. SUMOylation is one of many types of modification that can change and regulate the function of a protein. Such modifications mainly involve the attachment of small groups, for example phosphate, acetyl or methyl groups, sugars or lipids. Alternatively, a protein can be modified through the covalent attachment of another, usually smaller, protein. The best-known such modifier is ubiquitin, whose main function is to 'tag' other proteins for degradation. Since ubiquitin's discovery in the late 1970s, other ubiquitin-related modifiers have been identified, of which SUMO seems to affect the widest range of proteins. Although SUMO resembles ubiquitin in its three-dimensional structure and the way it links (conjugates) to other proteins, the functional consequences of SUMO conjugation are very different.

How and when was it discovered?

SUMO was discovered nearly two decades after ubiquitin, a delay mainly due to the fact that its attachment to other proteins is unstable and is lost with most experimental treatments. So it was work on an unusually stable modified protein, RanGAP1, that led to the discovery of SUMOylation in the mid-1990s. Refinement of experimental methods that prevent SUMO deconjugation soon led to the identification of hundreds of other SUMO targets.

Do all organisms have SUMOs?

SUMO is expressed by all eukaryotes (organisms such as yeast, plants and animals), but is absent from bacteria and from single-celled organisms known as archaea. Less complex eukaryotes (yeast, worms and flies) have a single SUMO gene, whereas plants and vertebrates express several SUMO proteins. In vertebrates, these can be divided into two sub-families: SUMO1 proteins and SUMO2/3 proteins. Although members of each sub-family are nearly identical, SUMO1 and SUMO2/3 proteins share only about 50% amino-acid sequence identity.

What is known about mammalian SUMO proteins?

Mammals express at least three SUMO proteins: SUMO1 and the twins SUMO2 and SUMO3. A fourth gene, coding for SUMO4, occurs in the human genome, but it is not clear whether its product can conjugate to other proteins. SUMO1, 2 and 3 seem to be expressed in all tissues and at all developmental stages, whereas SUMO4 expression seems to be restricted to the kidney and spleen. SUMO2/3 proteins are present at much higher levels than SUMO1. And although SUMO1 and SUMO2/3 use the same basic conjugation machinery, their functions are different. For example, they conjugate to different target proteins, are differentially affected by isopeptidase enzymes, respond differently to stress, and can be distinguished by their ability (SUMO2/3) or inability (SUMO1) to form SUMO chains. Furthermore, mouse embryos lacking SUMO1 die, indicating that SUMO2/3 cannot compensate for SUMO1 deficiency.

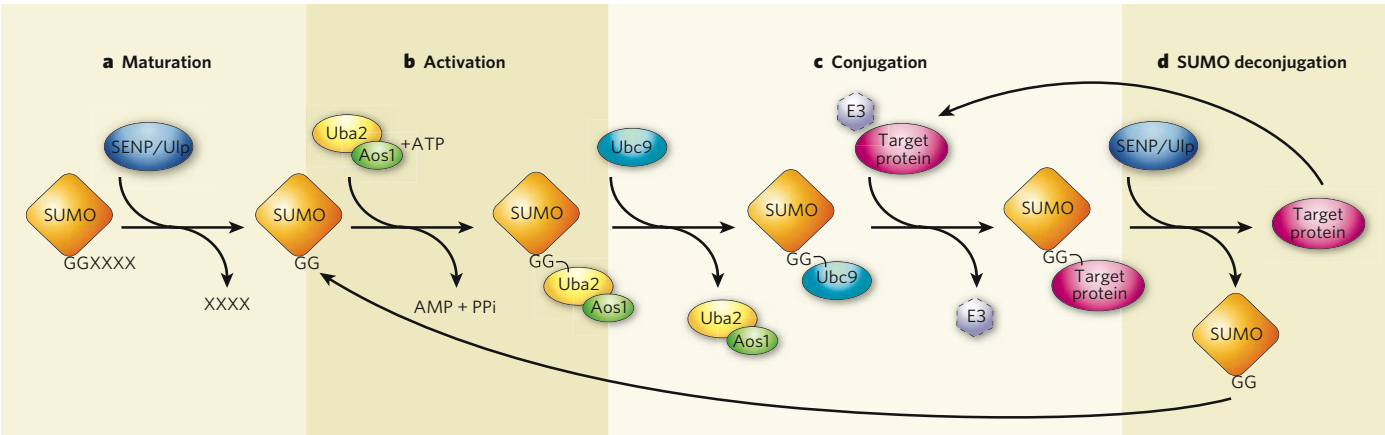


Figure 1 | SUMO conjugation and deconjugation. **a**, Newly synthesized SUMO is 'immature' as it cannot conjugate to its targets until two glycine residues (GG) close to its carboxy terminus are exposed in a reaction that removes some carboxy-terminus residues (XXXX) and is mediated by SUMO-specific proteolytic enzymes — SENP proteases in mammals and Ulp proteases in yeast. **b**, Mature SUMO is then activated in an energy (ATP)-consuming step by the E1 activating enzyme Aos1–Uba2. **c**, Subsequently, SUMO is transferred to the E2 conjugating enzyme Ubc9. In the last step, the carboxyl group of the glycine residue at SUMO's carboxy terminus forms an isopeptide bond with the amino group of a lysine residue in its target protein. This step is usually facilitated by E3 ligases, but some targets are efficiently SUMOylated by E2 alone. **d**, SUMOylation is reversible, because the SUMO-specific proteases of the SENP/Ulp family also efficiently cleave the isopeptide bond between SUMO and its target. Both the released SUMO and the target protein then become available for subsequent rounds of modification.

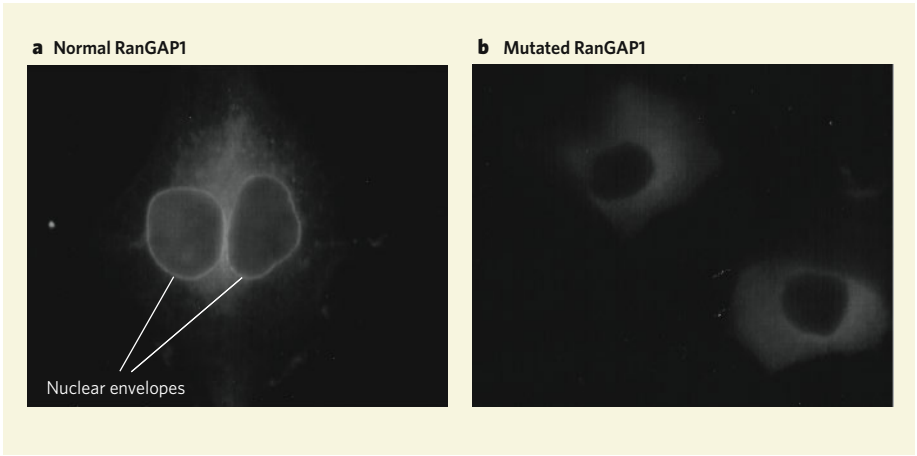


Figure 2 | SUMOylation and localization of target proteins. **a**, In the example shown, the SUMOylated RanGAP1 protein binds to a component of the nuclear-pore complex and is seen at the rim of the nuclei (nuclear envelopes) of two adjacent cells (grey). **b**, If this protein's lysine residue, to which SUMO covalently attaches, is replaced with an arginine residue — so that the mutated protein can no longer be SUMOylated — localization of RanGAP to the nuclear envelope is lost. (From R. Mahajan *et al. J. Cell Biol.* 140, 259–270; 1998).

How does SUMO conjugate to other proteins?

Like ubiquitin, SUMO can become covalently attached to a protein through an energy-consuming reaction cascade that requires the consecutive action of up to three enzymes: an E1 activating enzyme (the heterodimer Aos1–Uba2), an E2 conjugating enzyme (Ubc9) and one of several SUMO E3 ligases (Fig. 1). SUMO reacts with the free amino group (NH_2) of a lysine amino-acid residue on its target protein to form an isopeptide bond. Although SUMOylation usually results in the attachment of single SUMO moieties to one or a few acceptor lysine residues in the target protein, in some cases SUMO chains can form. This requires SUMO–SUMO isopeptide bond formation, which has been observed for SUMO in yeast and for SUMO2/3 in mammalian cells. Experiments in yeast suggest that SUMO chain formation is not essential for an organism's growth, but is required for spore formation in meiotic cell division.

Is this process reversible?

Yes. Many proteins undergo rapid cycles of SUMOylation and SUMO deconjugation; sometimes these occur only at certain times (for example, at a certain stage of the cell cycle) and places (such as the nucleus). Specific isopeptidases mediate SUMO deconjugation. These enzymes, of which two are known in yeast and six in mammals, are also involved in the maturation of newly synthesized SUMO (Fig. 1a, d). Very few proteins seem to be stably conjugated with SUMO; one such protein, RanGAP1, is protected from SUMO deconjugation through its association with a second protein.

What are the effects of SUMOylation?

These cannot be predicted easily, as they depend on the function of the target protein. SUMOylation could, for example, affect the target protein's intracellular localization (Fig. 2), interactions, stability and activity. At the molecular level, these highly variable outcomes can occur through one of three mechanisms (Fig. 3). First, SUMO can mask a binding site in its target, thereby inhibiting the target protein's interaction with other proteins. Second, SUMO can increase the number of binding sites on its target (after all, it contains around 100 amino acids), thus facilitating the binding of other molecules such as proteins or DNA. Finally, if the SUMOylated protein contains a second, non-covalent binding site for SUMO, a change in its conformation, and thus activity, can be induced.

How does SUMO interact with other proteins?

A SUMOylated protein can undergo specific interactions if its downstream effectors have low affinity for both its non-SUMOylated form and free SUMO. This is because SUMOylation can generate a binary binding site, allowing high-affinity interactions. Only one class of SUMO-interacting motif (SIM) has been identified. These motifs, which usually consist of a hydrophobic core flanked by acidic amino acid residues, mediate low-affinity, non-covalent interactions between SIM-containing proteins and free SUMO. SIMs have been found in the enzymes of the SUMO conjugation machinery, in SUMO target proteins and in proteins involved in SUMO-dependent repression of gene transcription.

What cellular processes are affected by SUMOylation?

Initially it was thought that SUMOylation was mainly involved in activities in the nucleus, such as transport between the nucleus and the cytoplasm, gene transcription, DNA replication and repair, chromosome segregation, and homologous recombination — the rearrangement of DNA strands during cell division. But with the identification of an increasing number of SUMO targets, it now seems that SUMOylation affects many fundamental pathways in both the nucleus and the cytoplasm (Fig. 4). These processes encompass basic metabolism, ion transport, fission and fusion of the energy-generating organelles known as mitochondria, and protein transport along the axons of neurons. SUMOylation can thus be added to the short-list of commonly used reversible modifications that includes phosphorylation, acetylation and ubiquitination.

Are SUMO modifications essential?

In most eukaryotes, reversible SUMOylation seems to be essential. In budding yeast, SUMO, the E1 and E2 enzymes and an isopeptidase are all necessary for survival. Similarly, depletion of the E2 enzyme in worms (such as *Caenorhabditis elegans*), plants (*Arabidopsis thaliana*) and mice results in embryonic death. A rare exception is fission yeast, in which disruption of its single SUMO gene results in viable but abnormal cells that have severe growth defects and aberrant cell division. Less clear is the role of individual SUMO proteins in organisms that express more than one SUMO. Mammalian SUMO1 seems to be essential, but mice lacking genes encoding SUMO2 and SUMO3 have not been described. In

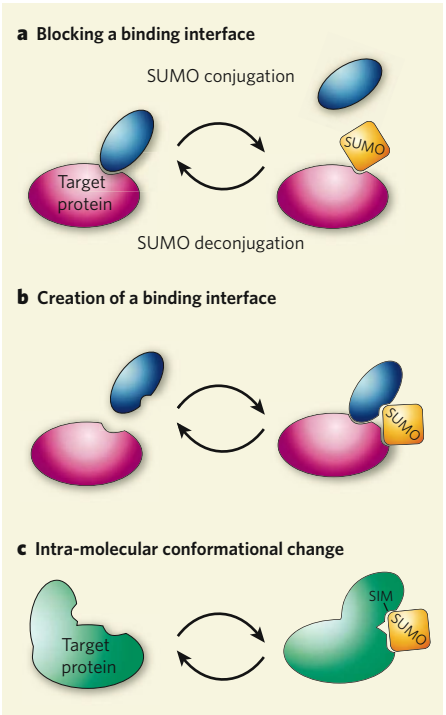


Figure 3 | Molecular switches. SUMO affects protein function in three ways. **a**, SUMOylation within or close to a target protein's interaction surface can prevent it from binding to another protein (blue). **b**, It can create a new binding surface. **c**, It can induce a conformational change if the target contains a SUMO-interaction motif (SIM) in addition to a SUMO acceptor site.

A. thaliana, disruption of the SUM1 or SUM2 gene does not affect survival, but simultaneous deletion of SUM1 and SUM2 genes is lethal to embryos. The extent to which specific SUMO proteins have non-essential or overlapping functions is not known.

Is there a disease connection?

There are some links between SUMOylation and disease, but not many. One example is the finding that a patient with a cleft lip and palate had a mutated copy of the gene encoding SUMO1 and thus reduced levels of the SUMO1 protein. Experiments in mice also indicate that reduced SUMO1 levels can cause developmental defects. Moreover, correlative evidence implicates misregulated SUMOylation in tumorigenesis: whereas the E2 conjugating enzyme is overexpressed in some human malignancies, a specific SUMO isopeptidase is overexpressed in others. It has also been suggested that SUMOylation is linked to neurodegenerative diseases, as it contributes to the toxicity of a pathogenic form of the huntingtin protein (involved in Huntington's disease) and modifies proteins implicated in Alzheimer's and Parkinson's diseases. Furthermore, some DNA viruses seem to exploit SUMOylation, either by blocking this process in the host cell or by hijacking SUMOylation for their own use. Therefore, as SUMOylation seems to have such dramatic cellular and physiological effects, it must be tightly regulated both at the level of individual proteins and throughout the cell.

How is the SUMOylation of individual proteins regulated?

This varies. For example, some proteins are SUMOylated only in response to DNA damage, and others conjugate with SUMO at specific stages of the cell cycle. Factors influencing SUMOylation of a protein include its cellular localization, the presence of specific stimuli, and the availability and activity of SUMO conjugation enzymes. Chemical modifications to the target protein can expose or mask a SUMO attachment site. Indeed, phosphorylation of a target protein near its SUMO attachment site can enhance or inhibit SUMOylation, whereas acetylation of the acceptor lysine residue prevents SUMOylation.

And throughout the cell?

SUMOylation levels can be increased or decreased throughout a cell or even a tissue (Fig. 4). In cells grown in culture, various stresses, such as heat shock and osmotic stress, can increase SUMOylation, whereas moderate levels of oxidative stress inactivate SUMO conjugation enzymes. SUMOylation is also inhibited in viral infection. Plants respond to stress induced by high salt concentration through markedly increasing SUMOylation of cellular proteins, and, in squirrels, SUMOylation in the brain, liver and kidney is greatly increased during hibernation. The underlying molecular mechanisms and physiological effects of these global changes are largely unknown. But by and large, SUMOylation seems to be correlated with protection against various forms of stress.

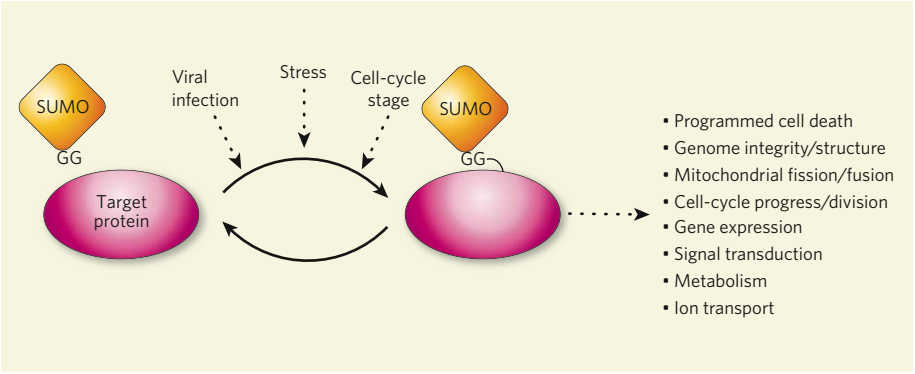


Figure 4 | SUMOylation is influenced by and affects a plethora of pathways. SUMO modifications are highly dynamic, and can be increased or decreased by many internal and external factors. In turn, SUMOylation can either increase or decrease the activity of its substrates. These can be components of a wide variety of cellular pathways.

How do viruses exploit the SUMOylation system?

To create an optimal environment for their propagation, viruses have evolved many ways to exploit SUMOylation. Several viral proteins — especially the first proteins generated after DNA virus infection, which replicate in the nucleus — rely on SUMOylation for their localization and/or function. Other viral proteins, including those of herpes viruses and a protein from CELO virus, can dramatically impair host-cell SUMOylation. In the case of CELO virus, the viral protein Gam1 targets the SUMO E1 enzyme for degradation by recruiting a ubiquitin E3 ligase, abolishing SUMOylation throughout the cell.

Can SUMOylation be predicted?

Unfortunately, not accurately, as the power of algorithms based on known SUMOylation sites is still rather limited. Many SUMO targets contain a 'consensus motif' of four amino acids that is recognized by the E2 enzyme Ubc9. But this motif also occurs in many proteins that are not SUMO targets. Moreover, many targets are modified at sites other than the consensus motif through various mechanisms. For example, the E2 conjugating enzyme carrying a SUMO can be recruited by means of a SIM in the target.

What are the latest advances in this field?

When SUMO was identified as a ubiquitin-related modifier, an obvious question was whether it would mark proteins for degradation as ubiquitin does. But one of the earliest identified SUMO targets, Iκ-Bα, is protected from degradation by SUMOylation, leading to the general belief that SUMO prevents degradation. This view must now be revised, as SUMO and ubiquitin can apparently join forces to mediate the degradation of a particular subset of proteins. Two different types of E3 ligase seem to bind to their targets only if they are SUMOylated. One example is the mammalian RNF4 ligase (and its homologues in yeast), which seems to recognize largely unknown targets only if they are SUMOylated. Another example is VHL, which seems to recognize the oxygen-sensitive transcription activator HIF1α under conditions of oxygen shortage only if HIF1α is SUMOylated.

And remaining questions are?

With the realization that SUMOylation affects thousands of proteins and virtually every cellular pathway in eukaryotes, our understanding of the molecular and physiological effects of SUMOylation at a cellular level and at the level of individual proteins clearly leaves much to be uncovered. Biochemical and cell-biological characterization of isopeptidases and E3 ligases has only begun, and other enzymes involved in SUMO conjugation and deconjugation, as well as downstream binding partners of SUMOylated proteins, will almost certainly be identified. Finally, we know little about the specific roles of SUMOylation during development, in adult organisms and in disease. ■

Erik Meulmeester and Frauke Melchior are in the Department of Biochemistry I, Faculty of Medicine, University of Göttingen, Humboldt-Allee 23, 37073 Göttingen, Germany.
e-mails: emeulme@gwdg.de; f.melchior@medizin.uni-goettingen.de

FURTHER READING
Johnson, E. S. *Annu. Rev. Biochem.* **73,** 355–382 (2004).
Hay, R. T. *Mol. Cell* **18,** 1–12 (2005).
Boggio, R. & Chiocca, S. *Curr. Opin. Microbiol.* **9,** 430–436 (2006).
Kerscher, O., Felberbaum, R. & Hochstrasser, M. *Annu. Rev. Cell Dev. Biol.* **22,** 159–180 (2006).
Mukhopadhyay, D. & Dasso, M. *Trends Biochem. Sci.* **32,** 286–295 (2007).
Martin, S., Wilkinson, K. A., Nishimune, A. & Henley, J. M. *Nature Rev. Neurosci.* **8,** 948–959 (2007).
Miura, K., Jin, J. B. & Hasegawa, P. M. *Curr. Opin. Plant Biol.* **10,** 495–502 (2007).
Geiss-Friedlander, R. & Melchior, F. *Nature Rev. Mol. Cell Biol.* **8,** 947–956 (2007).

ARTICLES

Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1

Janghoo Lim[1], Juan Crespo-Barreto[2], Paymaan Jafar-Nejad[1], Aaron B. Bowman[1]†, Ronald Richman[4], David E. Hill[5], Harry T. Orr[6] & Huda Y. Zoghbi[1,2,3,4]

Spinocerebellar ataxia type 1 (SCA1) is a dominantly inherited neurodegenerative disease caused by expansion of a glutamine-encoding repeat in ataxin 1 (*ATXN1*). In all known polyglutamine diseases, the glutamine expansion confers toxic functions onto the protein; however, the mechanism by which this occurs remains enigmatic, in light of the fact that the mutant protein apparently maintains interactions with its usual partners. Here we show that the expanded polyglutamine tract differentially affects the function of the host protein in the context of different endogenous protein complexes. Polyglutamine expansion in ATXN1 favours the formation of a particular protein complex containing RBM17, contributing to SCA1 neuropathology by means of a gain-of-function mechanism. Concomitantly, polyglutamine expansion attenuates the formation and function of another protein complex containing ATXN1 and capicua, contributing to SCA1 through a partial loss-of-function mechanism. This model provides mechanistic insight into the molecular pathogenesis of SCA1 as well as other polyglutamine diseases.

Expansion of an unstable translated CAG repeat located in different disease genes has so far been identified as causing nine dominantly inherited neurodegenerative disorders, the so-called polyglutamine diseases: Huntington's disease, spinobulbar muscular atrophy, dentatorubropallidoluysian atrophy, and six autosomal-dominant spinocerebellar ataxias[1]. As would be expected for dominant mutations, polyglutamine expansions confer toxic properties on the host proteins[1–3]; animal models genetically lacking the polyglutamine-containing proteins do not develop neurodegeneration[4–7]. However, expansion of the polyglutamine tract is necessary, but not sufficient, to cause pathology: in the case of SCA1, for example, expanded ataxin 1 (ATXN1) does not produce cerebellar degeneration if it lacks the nuclear localization signal[8] or the AXH (ATXN1-HBP1) domain[9], or if a serine to alanine substitution prevents phosphorylation at residue 776 (ref. 10). These and other studies in spinobulbar muscular atrophy and Huntington's disease indicate that protein domains outside of the polyglutamine tract have an important role in the selective neurotoxicity observed in these diseases[11–18]. Moreover, they suggest that there is a relationship between the normal functions of the wild-type proteins and the toxic functions of their expanded counterparts. Given that mouse and fly models overexpressing wild-type ATXN1 develop a mild version of SCA1 (ref. 19) begs the question of whether the glutamine expansion enhances some interactions to mediate the gain of function.

In an effort to address this question, we sought to characterize protein partners of ATXN1 that interact with it in a manner dependent on two criteria necessary for toxicity: polyglutamine expansion and phosphorylation at serine 776 (S776). We have identified RBM17 (RNA-binding motif protein 17) as a protein that meets these criteria. Here we show that ATXN1 forms at least two distinct, large native complexes, one containing capicua (CIC) and the other containing RBM17. Polyglutamine expansion alters the proportion of the mutant protein participating in the formation of these complexes *in vivo*. Furthermore, we show that the ATXN1–RBM17 complex causes disease by a gain-of-function mechanism, whereas the ATXN1–CIC complex causes a loss of function. These data not only show that there is an unexpected loss of function that contributes to SCA1 pathology, but provide a molecular mechanism to explain how glutamine expansion causes both gain and loss of function.

RBM17 preferentially interacts with glutamine-expanded ATXN1

To identify proteins that preferentially interact with ATXN1 in a manner dependent on both polyglutamine expansion and S776 phosphorylation, we performed a yeast two-hybrid screen[20] using several human ATXN1 constructs with either wild-type (30Q) or expanded (82Q) polyglutamine tracts and varying phosphorylation status at S776 (wild-type (S776), phosphorylation-defective (S776A), or phosphorylation-mimicked (S776D)) (Fig. 1a). This screen yielded a protein called RBM17 that binds specifically to ATXN1(S776D) (Fig. 1b). We verified the ATXN1–RBM17 interaction in mammalian cells by co-affinity purification assays using lysates from HEK293T cells transfected with glutathione *S*-transferase (GST)-tagged ATXN1 and Myc-tagged RBM17 (Fig. 1c, arrow). We also performed co-immunoprecipitation assays on mouse cerebellar extracts using a specific RBM17 antibody (Supplementary Fig. 1). The anti-RBM17 antibody co-immunoprecipitated Atxn1 from cerebellar extracts (Fig. 1d, arrow), suggesting that wild-type ATXN1 and RBM17 do indeed interact *in vivo*.

[1]Department of Molecular and Human Genetics, [2]Interdepartmental program in Cell and Molecular Biology, [3]Departments of Pediatrics and Neuroscience, [4]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA. [5]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]Institute of Human Genetics, Department of Biochemistry, Biophysics and Molecular Biology, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota 55455, USA. †Present address: Department of Neurology, Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, Tennessee 37232, USA.

To determine whether the interaction between RBM17 and ATXN1 depends on phosphorylation of S776 of ATXN1 in mammalian cells, we transfected Myc–RBM17 with GST–ATXN1(S776), GST–ATXN1(S776A), or GST–ATXN1(S776D) into HEK293T cells. Both wild-type ATXN1 and ATXN1(S776D) were able to precipitate RBM17, but ATXN1(S776A) showed almost no interaction with RBM17 (Fig. 2a, arrow, and Supplementary Fig. 2), indicating that the interaction of RBM17 with ATXN1 requires S776 phosphorylation. That RBM17 interacts more robustly with ATXN1(S776D) than with ATXN1(S776) suggests that S776 phosphorylation is important for the interaction. We next determined the domains responsible for the interaction between the two proteins using a series of ATXN1 and RBM17 deletion constructs. We found that the carboxy-terminal region of RBM17 interacts with the C-terminal sequence of ATXN1 harbouring the phosphorylated S776 residue (Supplementary Fig. 3).

To test the effect of polyglutamine tract length on the ATXN1–RBM17 interaction, we performed co-affinity purification assays by transfecting HEK293T cells with GST-fused ATXN1 containing 2Q, 30Q or 82Q. The interaction between Myc–RBM17 and GST–ATXN1 was strongly enhanced by expansion of the polyglutamine tract (Fig. 2b, arrow; see also Supplementary Fig. 4). We checked the phosphorylation level of S776 in ATXN1 containing a different polyglutamine tract length and found little if any enhancement in S776 phosphorylation of ATXN1 by polyglutamine expansion (Fig. 2b, bracket; see also Supplementary Fig. 5). These data suggest that expansion of the amino-terminal polyglutamine tract of ATXN1 alters the conformation of the C-terminal regions of the protein, making it more accessible to RBM17 binding.

We explored Rbm17 expression patterns in mouse brain and found Rbm17 protein abundant in Purkinje cell nuclei in wild-type mouse cerebella (Fig. 2c), consistent with the mRNA expression pattern shown in the Allen Brain Atlas[21]. We also tested Rbm17 expression in SCA1 transgenic mice (B05 line)[22], which express a polyglutamine-expanded form of human ATXN1 (ATXN1(82Q)) in Purkinje cells. Rbm17 expression remained high in the Purkinje cells of 18-week-old B05 mice, at a time when there is marked Purkinje cell atrophy (Fig. 2d and Supplementary Fig. 6), suggesting that Rbm17 is not depleted in SCA1.

To ascertain the relevance of the ATXN1–RBM17 interaction to ATXN1-induced neuropathology *in vivo*, we used a *Drosophila* model of SCA1 in which expression of polyglutamine-expanded human ATXN1 in the *Drosophila* eye causes retinal degeneration, ommatidial disorganization and fusion, and interommatidial bristle loss[19]. We established the physical interaction between ATXN1 and *Drosophila* RBM17 (dRBM17) (data not shown) and looked for genetic interaction between the two proteins by crossing SCA1 flies with *dRBM17* mutant flies (Fig. 3 and Supplementary Fig. 7). A heterozygous loss of one *dRBM17* allele partially suppressed ommatidial disorganization phenotypes of SCA1 flies at 30 °C (Fig. 3b), a temperature at which ATXN1(82Q) causes severe retinal degeneration (Fig. 3a). To verify the specificity of the genetic interaction, we generated two different lines of transgenic flies that expressed either human RBM17 (hRBM17) or dRBM17. Overexpression of either hRBM17 or dRBM17 alone in the *Drosophila* eye did not cause any obvious external morphological phenotypes at 25 °C (Supplementary Fig. 7). Co-expression of hRBM17 (or dRBM17) with ATXN1(82Q) worsened ommatidial disorganization and bristle loss in SCA1 flies (Fig. 3c, d). In contrast, hRBM17 (or dRBM17) had little effect when co-expressed with wild-type ATXN1(30Q) (Fig. 3e,

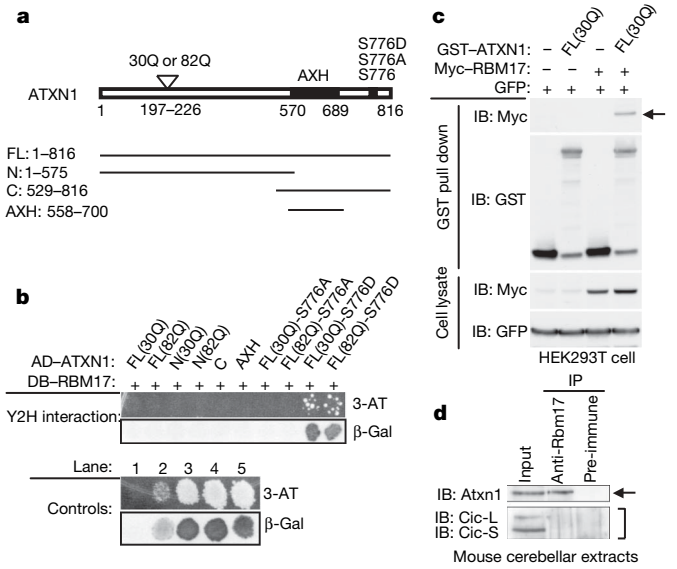


Figure 1 | ATXN1(S776D) but not ATXN1(S776A) interacts with RBM17. **a**, Schematic representation of the ATXN1 constructs. FL, full length. **b**, RBM17 specifically interacted with ATXN1(S776D) in the yeast two-hybrid screen. Activation domain (AD) and DNA binding domain (DB) of Gal4 were fused to human ATXN1 or RBM17, respectively. Yeast two-hybrid controls are: lane 1, negative control; lane 2, weak positive control; and lanes 3–5, strong positive controls. **c**, ATXN1 interacted with RBM17 in HEK293T cells by co-affinity purification assays. The top panel shows expression of Myc–RBM17 after affinity purification on glutathione-Sepharose 4B beads, demonstrating the ATXN1–RBM17 interaction (arrow). GST empty vector was used as a control (−). IB, Immunoblot. **d**, Co-immunoprecipitation of Atxn1 with Rbm17 from wild-type mouse cerebellar extracts. The anti-Rbm17 antibody co-immunoprecipitated Atxn1 (arrow), but not the long and short isoforms of capicua, Cic-L and Cic-S, respectively (bracket).

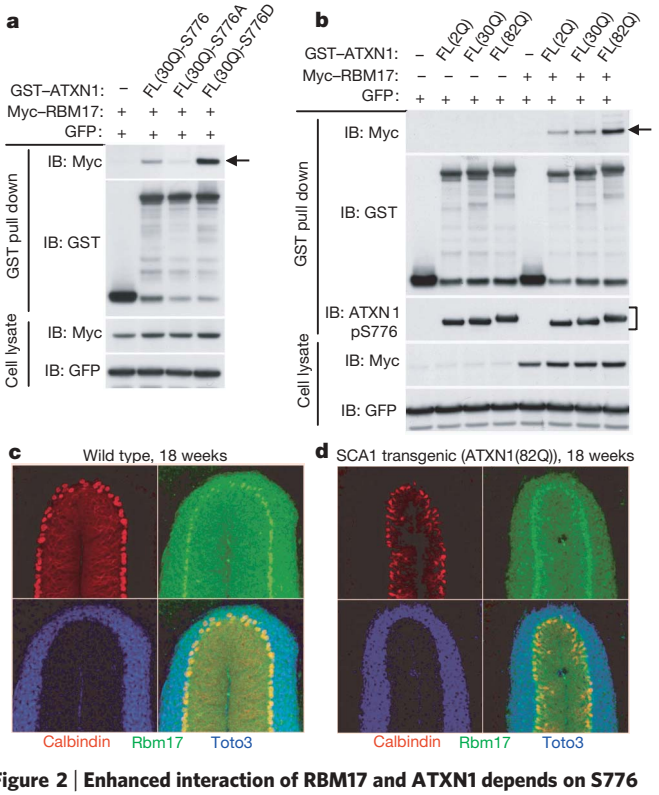


Figure 2 | Enhanced interaction of RBM17 and ATXN1 depends on S776 phosphorylation and polyglutamine tract expansion. **a**, Wild-type ATXN1 and ATXN1(S776D), but not ATXN1(S776A), interacted with RBM17. **b**, RBM17 bound most strongly to polyglutamine-expanded ATXN1. The bracket highlights that there is no obvious difference in S776 phosphorylation depending on polyglutamine expansion in ATXN1 when using a PN1168 antibody (specific to phosphorylated S776 of ATXN1) for immunoblotting. **c**, **d**, Rbm17 expression in the nuclei of Purkinje cells of 18-week-old mice from either wild type (FVB) (**c**) or SCA1 transgenic (ATXN1(82Q)) homozygotes (**d**). Rbm17 protein (green) was co-immunostained with calbindin (red, Purkinje cell marker) and Toto3 (blue, nuclear marker).

f). These genetic data suggest that RBM17 has a crucial role in mediating the toxicity of polyglutamine-expanded ATXN1 *in vivo*.

Native ATXN1 complexes contain RBM17

We next asked whether the ATXN1–RBM17 interaction is transient or whether the two proteins form a stable protein complex *in vivo*. We first analysed the elution profile of Rbm17 in wild-type mouse cerebellum using gel-filtration chromatography and compared it with the elution profile of Atxn1. Rbm17 eluted in a broad range of fractions (8 through 17) with at least two identifiable peaks (Fig. 4a, red and blue boxes): one peak coincided with large protein complexes with an estimated size larger than 4 MDa, at fraction 9, and a second peak occurred at fractions 12–13, consistent with small protein complexes at ~700 kDa in size. The elution profile of RBM17 in a human cell line, HEK293T cells, also revealed two elution peaks enriched for this protein in fractions similar to those found in wild-type mouse cerebellum, but relatively more RBM17 protein eluted in the large protein complexes detected in fraction 9 (Supplementary Fig. 8a, red and blue boxes). ATXN1 also eluted in a broad range of fractions and was associated with two isoforms of CIC (Fig. 4a and Supplementary Fig. 8a, top two rows), consistent with our previous work[23].

Given that the proteins showed broadly overlapping elution profiles, we sought to establish the fractions in which ATXN1 and RBM17 interact. We transfected HEK293T cells with Flag-tagged ATXN1(82Q), fractionated cell extracts, and immunoprecipitated Flag–ATXN1(82Q) from each of the fractions using anti-Flag antibody (Supplementary Fig. 8b). We found that ATXN1(82Q) immunoprecipitated endogenous RBM17 proteins from the large protein complexes (peak fraction 9, red box at the bottom panel), but not from the small protein complexes (peak fractions 12–13, blue box at the bottom panel). This suggests that ATXN1 and RBM17 interact in the large protein complexes. Incorporation of RBM17 and ATXN1 into large protein complexes was not caused by differential phosphorylation of ATXN1 at S776 (Supplementary Fig. 9). Of note, Atxn1 was not required for the presence of Rbm17 in the large protein complexes, as Rbm17 was still present in fractions 8–10 in *Atxn1* knockout ($Atxn1^{-/-}$) mouse cerebellum (Supplementary Fig. 10). This suggests that Rbm17 associates with other factors in the large protein complex in addition to forming complexes with Atxn1.

Glutamine expansion increases RBM17 in ATXN1 complexes

The findings that polyglutamine expansion enhanced the interaction of ATXN1 with RBM17 (Fig. 2b and Supplementary Fig. 4) and that ATXN1–RBM17 associated *in vivo* into large protein complexes (Supplementary Fig. 8b) suggested to us that more RBM17 molecules might incorporate into the large protein complexes in the presence of polyglutamine-expanded ATXN1. To test this hypothesis, we examined the elution profile of Rbm17 in cerebellar protein extracts from SCA1 knock-in mice that carry the polyglutamine-expanded (*Atxn1(154Q)*) allele[24] (Fig. 4b).

Rbm17 from SCA1 knock-in mouse cerebellar extracts fractionated into both large (Fig. 4b, red box) and small (Fig. 4b, blue box) protein complexes, but the relative ratio of Rbm17 incorporation into the large protein complexes was much greater in the SCA1 knock-in mice than in wild-type mice (Fig. 4). These data strongly support the hypothesis that more RBM17 is incorporated into large protein complexes because of enhanced interaction with polyglutamine-expanded ATXN1.

Two distinct large ATXN1 protein complexes

A major portion of wild-type Atxn1 co-elutes and forms large native protein complexes with two isoforms of Cic in mouse cerebellum[23] (Fig. 4a, top two rows). We therefore asked whether *in vivo* native protein complexes containing ATXN1–CIC contain RBM17 as well. Interestingly, despite the fact that both RBM17 and CIC interact with ATXN1, the two proteins were not detectable in the same ATXN1

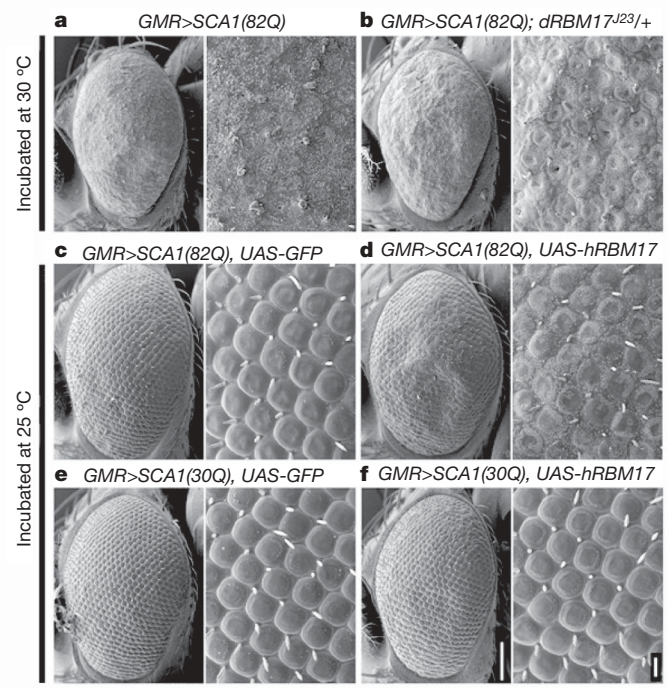


Figure 3 | RBM17 contributes to polyglutamine-expanded ATXN1 toxicity in the *Drosophila* eye. Scanning electron microscopy of adult *Drosophila* eyes. **a**–**f**, Loss of one *dRBM17* allele suppressed ATXN1(82Q)-mediated ommatidial disorganization (**a**, **b**), whereas overexpression of hRBM17 worsened abnormalities induced by ATXN1(82Q) but not by ATXN1(30Q) (**c**–**f**). Flies were raised at 30 °C (**a**, **b**) or 25 °C (**c**–**f**), and genotypes are: (**a**) *GMR-Gal4>UAS-ATXN1(82Q)*; (**b**) *GMR-Gal4>UAS-ATXN1(82Q); dRBM17^{J23}/+*; (**c**) *GMR-Gal4>UAS-ATXN1(82Q); UAS-GFP*; (**d**) *GMR-Gal4>UAS-ATXN1(82Q); UAS-hRBM17*; (**e**) *GMR-Gal4>UAS-ATXN1(30Q); UAS-GFP*; and (**f**) *GMR-Gal4>UAS-ATXN1(30Q); UAS-hRBM17*. Magnified images are on the right of each panel. Scale bars are 100 μm or 10 μm, respectively. Additional data with controls are available in Supplementary Fig. 7.

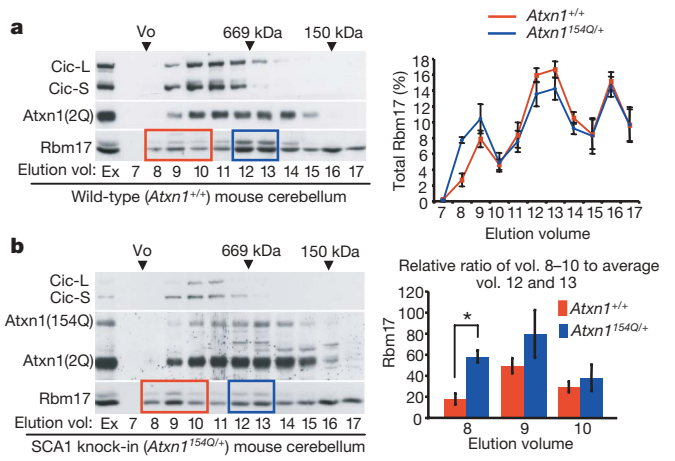


Figure 4 | Polyglutamine expansion enhances RBM17 incorporation into the large ATXN1 native protein complexes. **a**, **b**, Representative western blots of gel-filtration fractions of (**a**) wild-type ($Atxn1^{+/+}$) and (**b**) SCA1 knock-in ($Atxn1^{154Q/+}$) mouse cerebellar extracts analysed for Cic, Atxn1 and Rbm17. The right panel of **a** shows the elution profiles for Rbm17 plotted as the average percentage of protein (±standard error) in each fraction. The right panel of **b** shows the relative ratio of Rbm17 incorporation into large (red box, fractions 8–10) to small (blue box, fractions 12, 13) protein complexes and the marked increase in Rbm17 incorporation into large protein complexes in $Atxn1^{154Q/+}$ mice (asterisk, $P < 0.005$, $n = 4$ for $Atxn1^{+/+}$ and $n = 3$ for $Atxn1^{154Q/+}$). Ex, extract; Vo, column void volume.

complexes in either mouse cerebellum (Figs 1d and 5a) or in HEK293T cells (Supplementary Fig. 11). This suggests that there are at least two different ATXN1-associated large protein complexes *in vivo*: one containing ATXN1 and RBM17, the other containing ATXN1 and CIC. Interestingly, these two proteins, and possibly the protein complexes associated with them, have contrary effects on SCA1 pathology: extra RBM17 augments expanded ATXN1 toxicity in *Drosophila* (Fig. 3), but extra CIC represses it[23]. Consistent with these results as well as the accumulating evidence that nuclear inclusions serve a protective role in SCA1-affected cells[25], we observed that CIC enhanced, whereas RBM17 slightly suppressed, ATXN1 nuclear inclusion formation (Supplementary Fig. 12). Overexpression of Atxn1-like (a paralogue of Atxn1), which can interact with Atxn1 and Cic, also increases sequestration of mutant Atxn1 into inclusions in Purkinje cell nuclei of SCA1 knock-in mice[25].

The identification of two biochemically and functionally distinct large ATXN1 protein complexes led us to test whether RBM17 and CIC compete for interaction with ATXN1. We found in Neuro-2a as well as in HEK293T cells that RBM17 and CIC do indeed compete for ATXN1 interaction (Fig. 5b and Supplementary Fig. S11). These data, along with results described above, indicate that there are at least two distinct endogenous protein complexes associated with ATXN1.

SCA1 involves gain and partial loss of ATXN1 function

Next, we wished to address the question of how the RBM17- and CIC-containing endogenous protein complexes affect each other's formation and function, and how their interactions influence SCA1 neuropathology. Given that ATXN1–RBM17 interactions are polyglutamine-length-dependent and that RBM17 and CIC can compete with each other to form distinct ATXN1-containing protein complexes, we infer that polyglutamine expansion in ATXN1 favours the formation of the RBM17-containing protein complex, leaving mostly the wild-type ATXN1 to form a CIC-containing protein complex. Bearing in mind that the expression level of wild-type ATXN1 in SCA1 knock-in mice and in human patients is reduced to half the levels in healthy controls, and that Cic expression is also significantly reduced in SCA1 knock-in ($Atxn1^{154Q/+}$) mice (Supplementary Fig. 13), formation of expanded ATXN1–RBM17-containing complexes in soluble fractions is increased but the formation of wild-type ATXN1–CIC-containing protein complexes is decreased (Fig. 4). These considerations raise the interesting possibility that SCA1 is caused not only by a toxic gain-of-function mechanism but also by a partial loss of wild-type ATXN1 function.

To ascertain whether SCA1 neuropathology is affected by the presence or absence of the wild-type Atxn1 protein, we crossed *Atxn1* heterozygote ($Atxn1^{+/-}$) animals with the SCA1 knock-in ($Atxn1^{154Q/+}$) mice. $Atxn1^{154Q/-}$ mice showed a worsened performance on the rotarod than $Atxn1^{154Q/+}$ animals (Fig. 6a). Furthermore, $Atxn1^{154Q/-}$ mice had a shortened lifespan compared with $Atxn1^{154Q/+}$ animals (Fig. 6b). Given that *Atxn1* heterozygotes ($Atxn1^{+/-}$) have no discernable phenotype on their own, these data

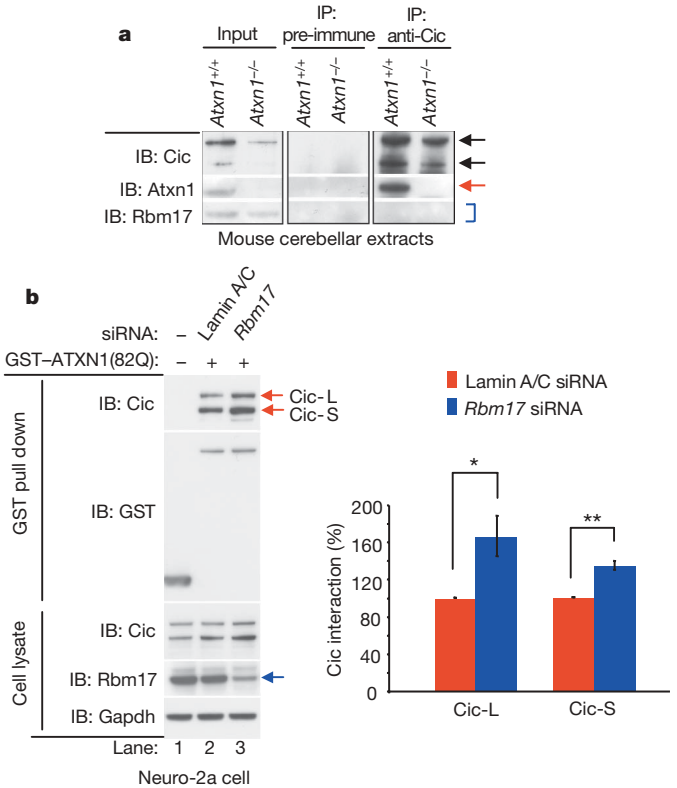


Figure 5 | Rbm17 and Cic form two distinct protein complexes that compete with each other. a, The Cic anti-serum co-immunoprecipitated Atxn1 (arrow) but not Rbm17 (bracket) from mouse cerebellar extracts. **b**, Atxn1 co-affinity purified more Cic protein when Rbm17 expression was reduced. The interaction of GST–ATXN1(82Q) with endogenous Cic was strongly increased (red arrows, compare lane 2 with lane 3) when endogenous Rbm17 expression was decreased (blue arrow) in Neuro-2a cells. The right panel shows the normalized levels of co-affinity-purified Cic-L and Cic-S. Mean relative levels (siRNA control (lamin A/C siRNA) = 100%) and standard error are shown ($n = 4$; asterisk, $P = 0.056$; double asterisk, $P < 0.006$).

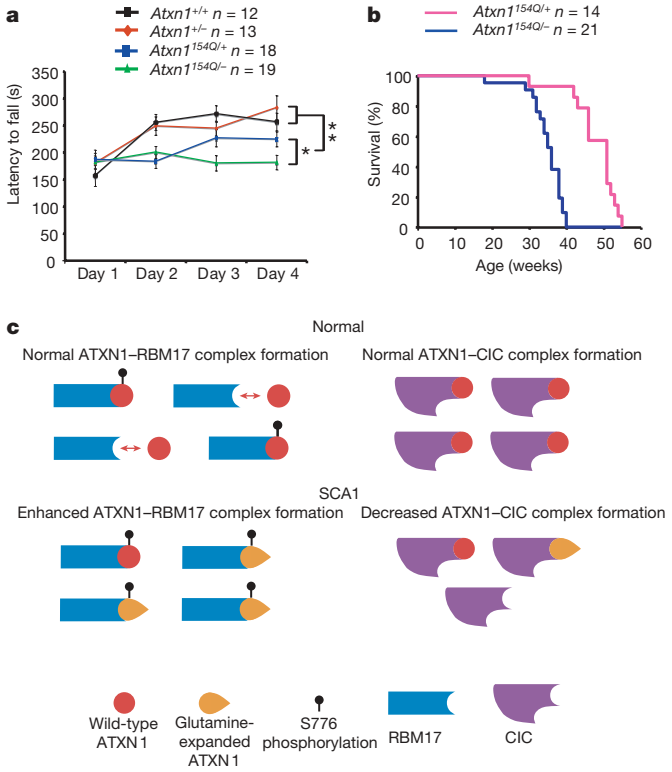


Figure 6 | Loss of wild-type Atxn1 function worsens SCA1 neuropathology in mice. a, b, $Atxn1^{154Q/-}$ animals showed a worsened rotarod performance (asterisk, $P < 0.05$; double asterisk, $P < 0.005$) and earlier lethality ($P < 1 \times 10^{-6}$) than $Atxn1^{154Q/+}$ mice. Error bar is the standard error of the mean (s.e.m.). **c**, Model for SCA1 neuropathology. In wild-type individuals, there are at least two distinct and mutually exclusive ATXN1-associated endogenous protein complexes *in vivo*. The formation of one of these complexes (ATXN1–RBM17) seems to be regulated in that it requires phosphorylation of ATXN1. Polyglutamine expansion in ATXN1 favours formation of the RBM17-containing complex, thereby enhancing one endogenous function and contributing to neuropathology by means of a gain-of-function mechanism. Polyglutamine expansion concomitantly decreases the formation of CIC-containing complex, resulting in a partial loss of function.

strongly suggest that wild-type Atxn1 and its interacting proteins are protective and that loss of the wild-type Atxn1 worsens SCA1 neuropathology.

Discussion

The results presented in this study provide a mechanistic explanation of how polyglutamine expansion can cause both gain and loss of normal protein function: the expansion differentially affects the interactions of the host protein in the context of distinct endogenous protein complexes *in vivo*. Previous work showed that polyglutamine-expanded ATXN1, like its wild-type counterpart, interacts with several nuclear proteins and incorporates into native complexes[23]. Some complexes, like those containing CIC, are relatively stable, whereas ATXN1's interactions with GFI-1 and TIP60-RORα are more transient[9,23,26]. RBM17 is unique among these factors, however, in that it is the only protein discovered so far whose interaction with ATXN1 is regulated by the length of the latter's polyglutamine tract and phosphorylation status at S776, two features that are absolute requirements for SCA1 pathogenesis.

Previously we have shown that wild-type and polyglutamine-expanded ATXN1 associate into large and small protein complexes (Fig. 4b, second row) and that incorporation of polyglutamine-expanded ATXN1 into the large complexes causes SCA1 neuropathology[23,25]. In this study, we found that there are at least two distinct large native protein complexes associated with ATXN1: one containing CIC and the other RBM17. Polyglutamine expansion in ATXN1 strongly enhanced the formation of an ATXN1–RBM17-containing protein complex in both cell culture and in SCA1 knock-in mouse cerebellum, providing a mechanistic explanation for the toxic gain of function. RBM17 is expressed in the nuclei of a broad range of neuronal and non-neuronal tissues[21,27]. It is interesting that both RBM17 (refs 28, 29) and ATXN1 (ref. 30) seem to be involved in regulation of RNA metabolism. One potential molecular function of this protein complex might thus be RNA splicing. This hypothesis receives some support from the observation that ATXN1 and RBM17 incorporation into large protein complexes was strongly decreased after RNase treatment (Supplementary Fig. 14).

Mice lacking wild-type Atxn1 but carrying polyglutamine-expanded Atxn1 suffered more severe disease and greater lethality, suggesting that either wild-type Atxn1 protects against the mutant protein or is essential for neuronal integrity. It is noteworthy that *Atxn1* null mice do not develop SCA1 features[4] and that polyglutamine-expanded ATXN1 does not differentially interact with wild-type ATXN1 (Supplementary Fig. 15), arguing against loss of function or dominant-negative mechanisms. Our data suggest that wild-type ATXN1 competes with polyglutamine-expanded ATXN1 to decrease formation of a toxic protein complex containing RBM17 (Supplementary Fig. 16). Nevertheless, we cannot rule out the possibility that ATXN1–CIC-containing complex is also neuroprotective. We found that glutamine-expanded ATXN1 preferentially forms a complex with RBM17, whereas wild-type ATXN1 is found in the CIC-containing complex. Therefore, when one allele is glutamine-expanded, the levels of wild-type ATXN1 decrease and less wild-type ATXN1–CIC complexes can be formed. Our previous findings[23] that the repressive activity of CIC is more robust in the presence of wild-type ATXN1 than with glutamine-expanded ATXN1 argue that a net effect of decreased ATXN1–CIC activity and the decreased co-repressive activity owing to mutant ATXN1 might also contribute to SCA1 pathogenesis. Although these data indicate that ATXN1–CIC complexes cause a partial loss of wild-type ATXN1 function, we cannot exclude the possibility that a mutant ATXN1–CIC complex might actively contribute to SCA1 pathogenesis. We propose a two-pronged model of SCA1 neurodegeneration in which augmented function of a particular stable endogenous protein complex is combined with a simultaneous loss of function of other, stable endogenous protein complexes (Fig. 6c).

This dual model may apply equally well to several other dominant neurodegenerative diseases caused by a gain-of-function mutational mechanism including other polyglutamine diseases[31–38], prion disease[39,40] and Alzheimer's disease[41,42]. Relatively recent work indicates that some of the polyglutamine diseases might also involve a partial loss of function of the wild-type host proteins[31–38]. However, the clear mechanism by which polyglutamine expansion causes both toxic gain of function and simultaneously loss of function of the disease-causing protein has not been clarified so far. In Huntington's disease, loss of mouse huntingtin (Htt) function intensifies neurodegeneration in transgenic models bearing the expanded protein[32–34]. Htt regulates REST (also called NRSF) activity differentially depending on whether it is wild type or polyglutamine-expanded[37]. Although the mechanism has not been fully described, the mutant Htt may exert a dominant deleterious effect on REST activity by losing its ability to sequester REST in the cytoplasm. In the dominantly inherited prion diseases, mutant prion proteins cause toxicity mainly through a toxic gain-of-function mechanism by interfering with essential cellular processes and activating cell death pathways[40]. In addition, some pathogenic mutations might also impair certain physiological functions of the normal host protein, contributing to disease by loss of a neuroprotective function of wild-type prion protein[39,40]. In the case of Alzheimer's disease, increased production and aggregation of the amyloid beta peptide by Alzheimer's-disease-related presenilin mutations is widely accepted to have a key role in Alzheimer's disease through a dominant gain-of-function mechanism[42]. However, recent work demonstrates that inactivation of presenilin in the adult cerebral cortex causes progressive memory loss and neurodegeneration that strikingly resemble Alzheimer's disease[41]. Whether presenilin mutations contribute to Alzheimer's disease through a gain- and/or a loss-of-function mechanism is not yet understood, but our finding that the interactions of the mutant protein with its usual partners are differentially affected by the polyglutamine expansion offers a mechanistic explanation for how mutant proteins can gain and lose function simultaneously.

METHODS SUMMARY

Yeast two-hybrid mating type screens[43] of human open reading frame (ORF) clones were performed using the hORFeome v1.1, as described[20]. The co-affinity purification and co-immunoprecipitation experiments were performed using cell lysates or mouse cerebellar extracts, as described[20,23]. Gel-filtration chromatography was performed using soluble extracts from about 20-week-old mouse cerebellum or cells as described[23]. *Drosophila* SCA1 model and *dRBM17* mutant flies were used to test a genetic interaction between the two proteins. The $dRBM17^{J23}$ allele is a strong hypomorphic or amorphic allele of *dRBM17* whereas *Df(2Lh)D1* is a deficiency that spans the *dRBM17* region[28]. Processing and image acquisition of *Drosophila* eye for scanning electron microscopy were performed as described[23]. Immunofluorescence staining and rotarod analysis were performed as described[24,25].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 November 2007; accepted 21 January 2008.
Published online 12 March 2008.

1. Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **30,** 575–621 (2007).
2. Harjes, P. & Wanker, E. E. The hunt for huntingtin function: interaction partners tell many different stories. *Trends Biochem. Sci.* **28,** 425–433 (2003).
3. Li, S. H. & Li, X. J. Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends Genet.* **20,** 146–154 (2004).
4. Matilla, A. *et al.* Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-pulse facilitation. *J. Neurosci.* **18,** 5508–5516 (1998).
5. Yeh, S. *et al.* Generation and characterization of androgen receptor knockout (ARKO) mice: an *in vivo* model for the study of androgen functions in selective tissues. *Proc. Natl Acad. Sci. USA* **99,** 13498–13503 (2002).
6. Zeitlin, S., Liu, J. P., Chapman, D. L., Papaioannou, V. E. & Efstratiadis, A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nature Genet.* **11,** 155–163 (1995).
7. Kiehl, T. R. *et al.* Generation and characterization of Sca2 (ataxin-2) knockout mice. *Biochem. Biophys. Res. Commun.* **339,** 17–24 (2006).

8. Klement, I. A. *et al.* Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* **95,** 41–53 (1998).
9. Tsuda, H. *et al.* The AXH domain of ataxin-1 mediates neurodegeneration through its interaction with Gfi-1/senseless proteins. *Cell* **122,** 633–644 (2005).
10. Emamian, E. S. *et al.* Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. *Neuron* **38,** 375–387 (2003).
11. McManamny, P. *et al.* A mouse model of spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* **11,** 2103–2111 (2002).
12. Katsuno, M. *et al.* Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. *Neuron* **35,** 843–854 (2002).
13. Chevalier-Larsen, E. S. *et al.* Castration restores function and neurofilament alterations of aged symptomatic males in a transgenic mouse model of spinal and bulbar muscular atrophy. *J. Neurosci.* **24,** 4778–4786 (2004).
14. Sopher, B. L. *et al.* Androgen receptor YAC transgenic mice recapitulate SBMA motor neuronopathy and implicate VEGF164 in the motor neuron degeneration. *Neuron* **41,** 687–699 (2004).
15. Graham, R. K. *et al.* Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell* **125,** 1179–1191 (2006).
16. Warby, S. C. *et al.* Huntingtin phosphorylation on serine 421 is significantly reduced in the striatum and by polyglutamine expansion *in vivo*. *Hum. Mol. Genet.* **14,** 1569–1577 (2005).
17. Luo, S., Vacher, C., Davies, J. E. & Rubinsztein, D. C. Cdk5 phosphorylation of huntingtin reduces its cleavage by caspases: implications for mutant huntingtin toxicity. *J. Cell Biol.* **169,** 647–656 (2005).
18. Steffan, J. S. *et al.* SUMO modification of Huntingtin and Huntington's disease pathology. *Science* **304,** 100–104 (2004).
19. Fernandez-Funez, P. *et al.* Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature* **408,** 101–106 (2000).
20. Lim, J. *et al.* A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* **125,** 801–814 (2006).
21. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–176 (2007).
22. Burright, E. N. *et al.* SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* **82,** 937–948 (1995).
23. Lam, Y. C. *et al.* ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell* **127,** 1335–1347 (2006).
24. Watase, K. *et al.* A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron* **34,** 905–919 (2002).
25. Bowman, A. B. *et al.* Duplication of Atxn1l suppresses SCA1 neuropathology by decreasing incorporation of polyglutamine-expanded ataxin-1 into native complexes. *Nature Genet.* **39,** 373–379 (2007).
26. Serra, H. G. *et al.* RORα-mediated Purkinje cell development determines disease severity in adult SCA1 mice. *Cell* **127,** 697–708 (2006).
27. Sampath, J. *et al.* Human SPF45, a splicing factor, has limited expression in normal tissues, is overexpressed in many tumors, and can confer a multidrug-resistant phenotype to cells. *Am. J. Pathol.* **163,** 1781–1790 (2003).
28. Chaouki, A. S. & Salz, H. K. *Drosophila* SPF45: A bifunctional protein with roles in both splicing and DNA repair. *PLoS Genet* **2,** e178 (2006).
29. Lallena, M. J., Chalmers, K. J., Llamazares, S., Lamond, A. I. & Valcarcel, J. Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. *Cell* **109,** 285–296 (2002).
30. Yue, S., Serra, H. G., Zoghbi, H. Y. & Orr, H. T. The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Hum. Mol. Genet.* **10,** 25–30 (2001).
31. Zuccato, C. *et al.* Widespread disruption of repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy at its target genes in Huntington's disease. *J. Neurosci.* **27,** 6972–6983 (2007).
32. Auerbach, W. *et al.* The HD mutation causes progressive lethal neurological disease in mice expressing reduced levels of huntingtin. *Hum. Mol. Genet.* **10,** 2515–2523 (2001).
33. Van Raamsdonk, J. M. *et al.* Loss of wild-type huntingtin influences motor dysfunction and survival in the YAC128 mouse model of Huntington disease. *Hum. Mol. Genet.* **14,** 1379–1392 (2005).
34. Leavitt, B. R. *et al.* Wild-type huntingtin reduces the cellular toxicity of mutant huntingtin *in vivo*. *Am. J. Hum. Genet.* **68,** 313–324 (2001).
35. Cattaneo, E., Zuccato, C. & Tartari, M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nature Rev. Neurosci.* **6,** 919–930 (2005).
36. Thomas, P. S. Jr *et al.* Loss of endogenous androgen receptor protein accelerates motor neuron degeneration and accentuates androgen insensitivity in a mouse model of X-linked spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* **15,** 2225–2238 (2006).
37. Zuccato, C. *et al.* Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genet.* **35,** 76–83 (2003).
38. Friedman, M. J. *et al.* Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nature Neurosci.* **10,** 1519–1528 (2007).
39. Li, A., Piccardo, P., Barmada, S. J., Ghetti, B. & Harris, D. A. Prion protein with an octapeptide insertion has impaired neuroprotective activity in transgenic mice. *EMBO J.* **26,** 2777–2785 (2007).
40. Harris, D. A. & True, H. L. New insights into prion structure and toxicity. *Neuron* **50,** 353–357 (2006).
41. Shen, J. & Kelleher, R. J. III. The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. *Proc. Natl Acad. Sci. USA* **104,** 403–409 (2007).
42. Van Broeck, B., Van Broeckhoven, C. & Kumar-Singh, S. Current insights into molecular mechanisms of Alzheimer disease and their implications for therapeutic approaches. *Neurodegener. Dis.* **4,** 349–365 (2007).
43. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437,** 1173–1178 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Vidal for the human ORFeome and yeast two-hybrid screening technology; J. Valcárcel and W. Perry for anti-RBM17 antibody; H. Salz for *dRBM17* mutant flies; Y. He for generating the RBM17 transgenic flies; H. Bellen and H. Jafar-Nejad, and members of the Zoghbi laboratory, for comments on the manuscript; and V. Brandt for editorial input. This research was supported by the NIH grants (H.Y.Z., H.T.O., M.V.), cores of the BCM-MRDDRC, the Ellison Foundation and the W.M. Keck Foundation awarded to M.V. and D.E.H., and Institute Sponsored Research from the DFCI Strategic Initiative in support of Center for Cancer Systems Biology. H.Y.Z. is an investigator with the Howard Hughes Medical Institute.

Author Contributions J.L. and H.Y.Z. designed the experiments. J.L. performed the majority of the experiments with the exception of the mouse genetic interaction in Fig. 6a, b (J.C.-B. and A.B.B.), and immunoprecipitation assays in Figs 1d and 5a (P.J.-N.). R.R. provided technical assistance performing gel-filtration chromatography. D.E.H. provided the ORFeome library and advice. Data analyses and interpretation were conducted by J.L., H.T.O. and H.Y.Z. J.L. and H.Y.Z. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.Y.Z. (hzoghbi@bcm.tmc.edu).

METHODS

Yeast two-hybrid screens. Yeast two-hybrid mating type screens[43] of human open reading frame (ORF) clones were performed using the hORFeome v1.1 as described[20]. For re-testing the ATXN1–RBM17 interaction in yeast, MaV103 (*MAT**a***) yeast strain transformed with various AD-ATXN1 constructs was individually mated against MaV203 (*MAT*α) containing DB-RBM17. Resulting interactions were tested on 3-AT plates (synthetic complete media plates lacking leucine, tryptophan and histidine and containing 20 mM 3-amino-2,4-triazole (3-AT)) as well as by β-Gal filter lift assay. Yeast two-hybrid controls are the following: lane 1 expresses AD and DB without any fusion, acting as a negative control. Lane 2 expresses AD-E2F1 and DB-pRB and shows a weak two-hybrid phenotype. Lanes 3, 4 and 5 express AD-Jun and DB-Fos, AD and Gal4p, and AD-E2F1 and DB-DP, respectively, and exhibit relatively strong two-hybrid phenotypes.

Co-affinity purification and co-immunoprecipitation assays. The co-affinity purification experiments were performed as described[20]. Briefly, HEK293T cells were plated the day before transfection at 1×10^5 cells per well in 6-well plates. The following day, cDNA constructs were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions and cells were cultured in DMEM medium with 10% fetal bovine serum. Two days later, cells were harvested and lysed with lysis buffer (0.5% NP-40, 20 mM Tris-HCl (pH 8.0), 150 to 180 mM NaCl, 1 mM EDTA, and complete protease inhibitor cocktail (Roche)) for 15 min on ice. After centrifugation, soluble protein complexes were purified using glutathione Sepharose 4B beads (Amersham), washed with lysis buffer, and analysed by SDS–PAGE and western blot. GFP expression was used as a transfection and loading control. The co-immunoprecipitation assays were performed using mouse cerebellar extracts as described[23].

Column fractionation and analysis. Gel-filtration (sizing) chromatography was performed using soluble extracts from about 20-week-old mouse cerebellum or HEK293T or Neuro-2a cells using the Amersham Pharmacia LCC-500 FPLC system as described[23]. Protein signal in each fraction (measured by densitometry) is divided by the total signal in all the fractions to determine the percentages. Thyroglobulin (669 kDa) and ADH (150 kDa) were used for gel-filtration standards. Monoclonal anti-HA agarose conjugate clone HA-7 (Sigma, A2095) and anti-Flag M2 affinity gel freezer-safe (Sigma, A2220) were used for immunoprecipitation after column fractionation.

Immunofluorescence and confocal microscopy. Immunofluorescence staining was performed as described[25]. Sections from frozen fixed cerebella of *Atxn1*$^{+/+}$, *Atxn1*$^{+/-}$, *Atxn1*$^{154Q/+}$, *Atxn1*$^{154Q/-}$ and SCA1 transgenic (*ATXN1(82Q)*) mice were co-stained with the following antibodies: rabbit anti-RBM17 (1:400), rabbit anti-ATXN1 (11NQ, 1:1,000) and mouse anti-calbindin (1:1,000; Sigma) antibodies. Fluorescent images were scanned using a Zeiss LSM510 confocal microscope and processed with ImageJ software and Adobe Photoshop.

Rotarod and statistical analyses. We performed *t*-tests (two-tailed, not assuming equal variances) and calculated standard error of the mean (s.e.m.) and 95% confidence intervals using Microsoft Excel. Rotarod analysis was performed as described previously[24] using 7-week-old animals. We analysed rotarod performance by repeated-measures ANOVA (ANOVA) using SPSS 11 software for Mac OSX.

***Drosophila* strains and scanning electron microscopy.** Full-length *hRBM17* and *dRBM17* cDNAs were cloned into the pUAST-dest vector using the Gateway system to generate *UAS-hRBM17* or *UAS-dRBM17* transgenic fly lines, respectively. The mutant and transgenic flies used in this study are *dRBM17*$^{\Delta 23}$, *Df(2Lh)D1* (ref. 28), *UAS-ATXN1(82Q)F7* and *UAS-ATXN1(30Q)F1* (ref. 19). Other strains were obtained from the Bloomington Stock Center (Flybase; http://www.flybase.org). For the genetic interaction, more than 50 adult flies per genotype were examined at day 2 after eclosion. Processing and image acquisition of *Drosophila* eye for scanning electron microscopy were performed as described[23].

ARTICLES

Clathrin is a key regulator of basolateral polarity

Sylvie Deborde[1], Emilie Perret[1]*, Diego Gravotta[1]*, Ami Deora[1], Susana Salvarezza[1], Ryan Schreiner[1] & Enrique Rodriguez-Boulan[1,2]

Clathrin-coated vesicles are vehicles for intracellular trafficking in all nucleated cells, from yeasts to humans. Many studies have demonstrated their essential roles in endocytosis and cellular signalling processes at the plasma membrane. By contrast, very few of their non-endocytic trafficking roles are known, the best characterized being the transport of hydrolases from the Golgi complex to the lysosome. Here we show that clathrin is required for polarity of the basolateral plasma membrane proteins in the epithelial cell line MDCK. Clathrin knockdown depolarized most basolateral proteins, by interfering with their biosynthetic delivery and recycling, but did not affect the polarity of apical proteins. Quantitative live imaging showed that chronic and acute clathrin knockdown selectively slowed down the exit of basolateral proteins from the Golgi complex, and promoted their mis-sorting into apical carrier vesicles. Our results demonstrate a broad requirement for clathrin in basolateral protein trafficking in epithelial cells.

Epithelial cells require a polarized distribution of their plasma membrane (PM) proteins to perform a variety of vectorial functions in absorption and secretion[1–3]. Generation of polarity requires mechanisms to sort the PM proteins into apical and basolateral domains, separated by tight junctions. Sorting is directed by sorting signals in the PM protein. Basolateral sorting signals are distinct determinants in the cytoplasmic domain; in some cases they resemble tyrosine and dileucine motifs similar to those used at the cell surface for clathrin-mediated receptor endocytosis[4–7]. An epithelial-specific adaptor, AP1B[8,9], localizes to recycling endosomes and mediates biosynthetic and/or post-endocytic sorting of basolateral PM proteins with tyrosine motifs (for example vesicular stomatitis virus (VSV) G protein (VSVG) and low-density lipoprotein receptor (LDLR)) and non-tyrosine motifs (for example transferrin receptor (TfR))[10–12]. AP1B is a tetrameric adaptor with clathrin-interacting domains in its γ subunit that co-localizes with clathrin-coated vesicles near the Golgi[13].

All of the above is compatible with a possible function of clathrin in the biosynthetic sorting and recycling of PM proteins. Indeed, an early cell-fractionation study detected a newly synthesized PM protein in clathrin-coated vesicles[14]. However, a requirement for clathrin in the biosynthetic route of PM proteins has never been shown. Acute crosslinking experiments failed to detect the involvement of clathrin in recycling of TfR to the PM of non-polarized cells[15], although more recent *in vitro* reconstitution experiments[16] and an RNA-mediated interference (RNAi) study[17] support the involvement of clathrin in non-polarized PM protein recycling. Here we study the involvement of clathrin in PM protein trafficking in polarized MDCK cells by using RNAi and crosslinking approaches to suppress clathrin expression or function, together with a battery of biochemical and live-imaging trafficking assays. Our experiments demonstrate a fundamental role of clathrin in PM protein transport to the basolateral membrane.

Knockdown of clathrin heavy chain in MDCK cells

In the epithelial cell line MDCK, as in other nucleated cells, clathrin localizes prominently to the Golgi region (Fig. 1a). Treatment of MDCK cells with two consecutive cycles (three days each) of siRNA against clathrin heavy chain (CHC) drastically decreased clathrin immunofluorescence in both perinuclear and PM pools (Fig. 1a). The Golgi apparatus had an overall normal shape and localization in clathrin-depleted cells. Western blot analysis (Fig. 1b) showed that the treatment depleted CHC (more than 90%) and clathrin light chains (CLCb, 45%; CLCa, 72%), the latter as a result of decreased stability on depletion of CHC, relative to control MDCK cells treated with siRNA against luciferase, as shown previously in non-polarized cells[18,19]. Clathrin-depleted MDCK cells grew more slowly than control cells (not shown), as reported previously for fibroblastic cells[20]. Clathrin depletion slowed the exit of the lysosomal hydrolase cathepsin D from the *trans*-Golgi network (TGN), as indicated by a delay in the processing of its TGN precursor procathepsin D (51 kDa) to endosomal (48 kDa) and lysosomal (34 kDa) forms of the enzyme[21] (Fig. 1c, d). Thus, clathrin depletion by siRNA was effective in abrogating the best-known non-endocytic trafficking function of clathrin, namely lysosomal hydrolase transport, in MDCK cells.

The development of epithelial polarity and most of our polarity assays depend crucially on the assembly of functional tight junctions and adherent junctions, which separate apical and basolateral PM domains[2,22]. Clathrin depletion did not prevent the formation of functional tight junctions. After three days of plating the cells at confluence, the transepithelial electrical resistance reached similar levels, about 100 $\Omega\,cm^2$, in both control and clathrin-depleted cells (Fig. 2a). The [^{3}H]inulin flux across the monolayer was similar in control and clathrin-depleted cells (Fig. 2b). Tight junctions had a normal morphology in clathrin-depleted cells, as determined by immunostaining of the tight-junction marker ZO1 (Fig. 2c). Par6, a polarity protein that localizes at tight junctions[23], was also found in its normal localization in clathrin-depleted cells (Supplementary Fig. 1). These experiments showed that clathrin-depleted MDCK cells develop normal tight junctions after three days at confluence.

[1]Department of Ophthalmology, Dyson Vision Research Institute, LC-300, and [2]Department of Cell and Developmental Biology, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10065, USA.
*These authors contributed equally to this work.

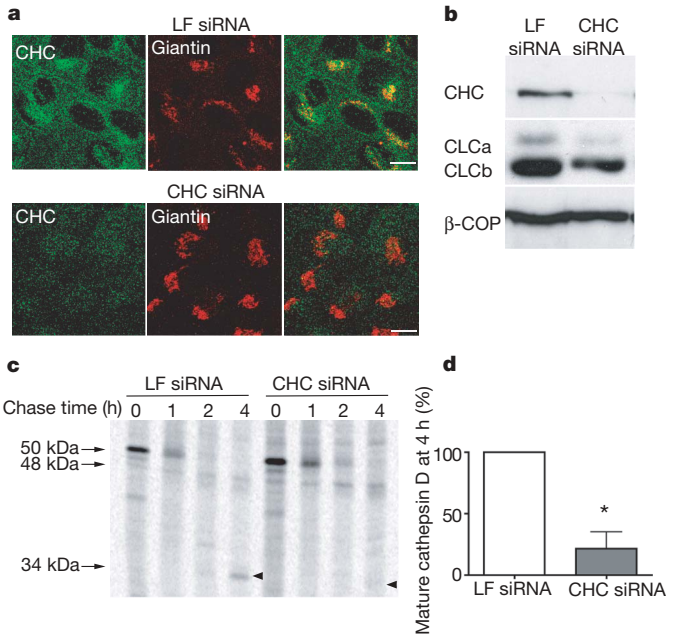


Figure 1 | Clathrin knockdown in MDCK cells. MDCK cells were electroporated twice with CHC or luciferase (LF) siRNAs and used three days later. **a**, Confocal fluorescence microscopy of monolayers stained with CHC antibody TD1 and Golgi marker giantin. CHC siRNA caused drastic depletion of perinuclear and PM clathrin pools (green); the Golgi complex (red) appeared slightly swollen but normal overall. Scale bar, 10 μm. **b**, Western blot of cells treated with CHC siRNA, demonstrating strong depletion of CHC (90%), CLCa (72%) and CLCb (45%). β-COP, β-coat protein. **c**, Impairment of maturation of procathepsin D into lower-molecular-weight forms by clathrin depletion (arrows, arrowheads). Pulse–chase and SDS–PAGE. **d**, Maturation of cathepsin D quantified by normalizing mature cathepsin D levels at $t = 4$ h over procathepsin D levels at $t = 0$. Mean and s.e.m. are shown ($P < 0.005$, $n = 3$).

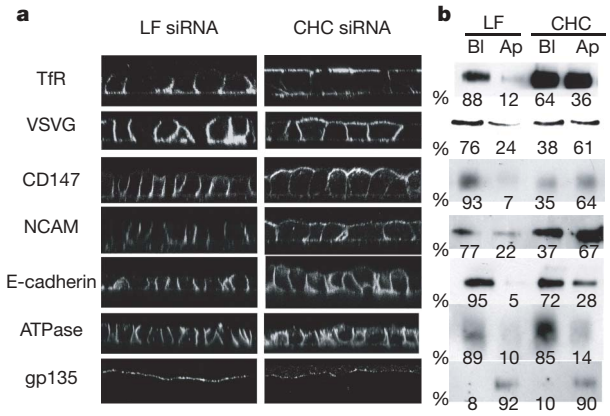


Figure 3 | Clathrin knockdown depolarizes basolateral proteins at steady state. **a**, Confocal microscopy of control and clathrin-depleted cells. **b**, Domain-selective biotinylation of control and clathrin-depleted cells. MDCK cells were confluent on filters for at least three days after treatment with luciferase (LF) or CHC siRNAs. Some markers were endogenous (TfR (**b**), E-cadherin, Na^+, K^+-ATPase β subunit (ATPase) and gp135). NCAM and CD147 were permanently transfected. VSVG and human TfR (**a**) were expressed by adenoviral infection. Bl, basolateral; Ap, apical.

Clathrin siRNA disrupts basolateral PM polarity

Next we studied the effect of clathrin depletion on the steady-state surface polarity of apical and basolateral PM markers in MDCK cells in control and clathrin-depleted MDCK cells that had been confluent for at least three days. Confocal microscopy (Fig. 3a) and domain-selective biotinylation (Fig. 3b) demonstrated a loss of polarity of most (five out of six) basolateral PM proteins tested, covering a broad range of basolateral signals. These include TfR (tyrosine-independent signal[24]); VSVG (tyrosine signal[25]); E-cadherin (dileucine motif[26]); neural cell adhesion molecule (NCAM; tyrosine-independent signal[27]) and CD147 (single leucine plus acidic cluster signal[28]). The polarity of Na^+, K^+-ATPase was not significantly affected, possibly reflecting a clathrin-independent sorting mechanism or its stabilization at the lateral PM by the ankyrin cytoskeleton[29]. The polarity of the apical PM markers gp135 (ref. 30) (Fig. 3a, b) and p75 (ref. 31) (Supplementary Fig. 2) was not affected by clathrin depletion.

Clathrin-depleted cells overexpressed basolateral PM proteins, which is consistent with clathrin's role in promoting trafficking of PM proteins to the degradative pathway[32]. However, quantification of the polarity of NCAM tagged with green fluorescent protein (GFP) in individual cells over a broad range of expression levels (Supplementary Fig. 3) showed that the disruption of its basolateral polarity could not be attributed to saturation of the basolateral sorting machinery[5].

Clathrin depletion disrupted the recycling of TfR and LDLR from recycling endosomes to the basolateral membrane, as expected from the involvement of AP1B in this route[10,12], and promoted the incorporation of both receptors into the apical recycling (transcytotic) route (Fig. 4a). Pulse–chase surface capture assays[11,28] demonstrated a large delay in the delivery of newly synthesized TfR, VSVG and CD147 to the basolateral PM and mis-sorting of the proteins to the apical PM (Fig. 4b). These experiments showed that clathrin depletion causes a specific defect in the transport and sorting of basolateral PM proteins in the biosynthetic and recycling routes.

Clathrin depletion slows Golgi exit of basolateral proteins

Sorting of PM proteins in epithelial cells takes place at a Golgi or post-Golgi level (see above). To determine whether clathrin knockdown

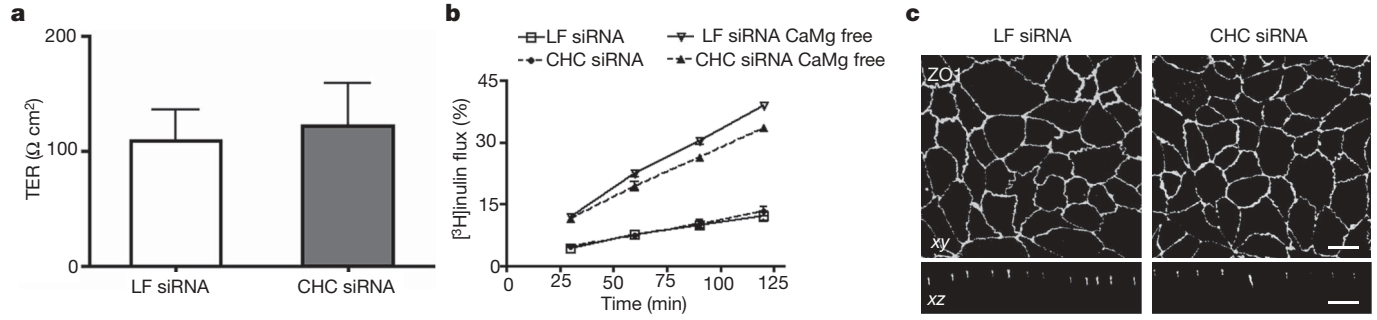


Figure 2 | Clathrin-depleted cells develop normal tight junctions. **a**, Development of normal transepithelial electrical resistance (TER) in MDCK cells treated with luciferase (LF) or CHC siRNAs, after three days at confluence. Means and s.e.m. are shown ($P = 0.45$, $n = 5$). **b**, Normal permeability to [^{3}H]inulin in clathrin-depleted MDCK cells. Means ± s.e.m. are shown for two experiments, with $n = 3$ for each experiment. **c**, Morphologically normal tight junctions in clathrin-depleted cells. Immunofluorescence with ZO1 antibody. Scale bar, 10 μm.

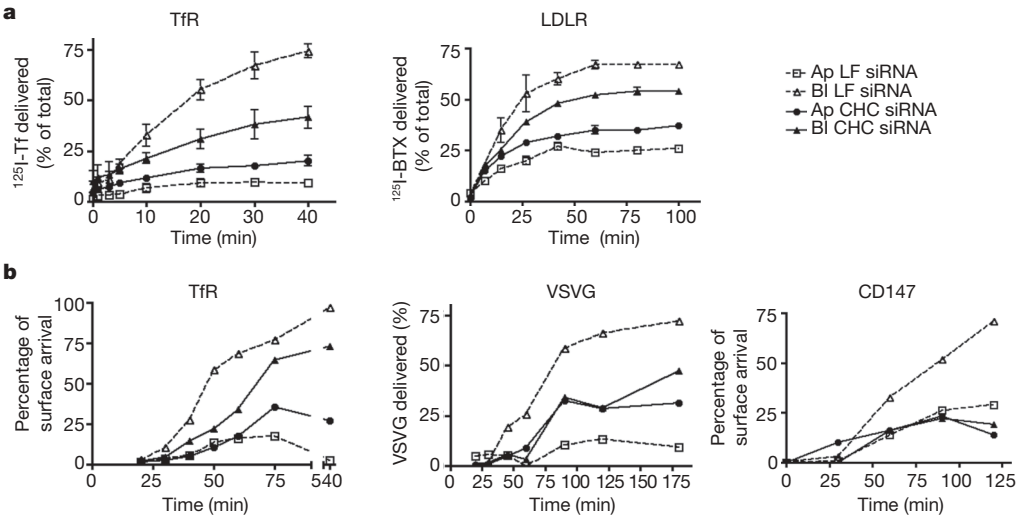


Figure 4 | Clathrin knockdown depolarizes surface delivery of basolateral proteins. a, Clathrin depletion decreases recycling accuracy of basolateral PM proteins. MDCK-cell monolayers on transwells expressing human TfR or LDLR fused to α-bungarotoxin (BTX) binding site (see Methods) were incubated for 2 h in the presence of ^{125}I-transferrin (Tf) (added basolaterally) or ^{125}I-BTX (added to both sides). Basolaterally released or apically released (transcytosed) ^{125}I-Tf was plotted as a fraction of the total endocytosed pool. Basolaterally or apically released ^{125}I-BTX was plotted as a fraction of the total endocytosed pool. Bl, basolateral; Ap, apical. Means ± s.e.m. are shown. **b**, Clathrin depletion causes retention and promotes depolarized delivery of newly synthesized basolateral PM proteins. TfR and VSVG were introduced by means of adenoviruses, whereas CD147 was permanently expressed. Cells were radiolabelled for 15 min and chased for different durations. Surface delivery of TfR at each time point is plotted as a percentage of total surface delivery at $t = 75$ min. Surface delivery of CD147 and VSVG is plotted as a percentage of total radioactively pulsed protein.

disrupted the trafficking of PM proteins at the Golgi complex, we introduced GFP-tagged basolateral (VSVG–GFP and NCAM–GFP) or apical (p75–GFP) PM proteins by adenovirus transduction (VSVG) or nuclear microinjection of their complementary DNAs (NCAM or p75) into subconfluent MDCK cells, promoted their accumulation at the TGN by incubation for 1–2 h at 20 °C, and then recorded, by live imaging, their exit from the TGN after transfer to 32 °C (refs 31, 33) (Fig. 5a–d). For VSVG–GFP, we used an MDCK cell line permanently expressing sialyl transferase tagged with monomeric red fluorescent protein (ST–RFP) to identify the position of the TGN and quantify the TGN-associated GFP (Fig. 5a). Clathrin knockdown markedly decreased the exit kinetics of both basolateral PM proteins, VSVG–GFP and NCAM–GFP ($t_{1/2}$ increased from 20–25 min to 70–75 min) but not the exit kinetics of apical PM protein p75 (Fig. 5b, right panel). VSVG–GFP remained strictly co-localized with ST–RFP in clathrin-depleted cells, suggesting that the transport block was at the TGN.

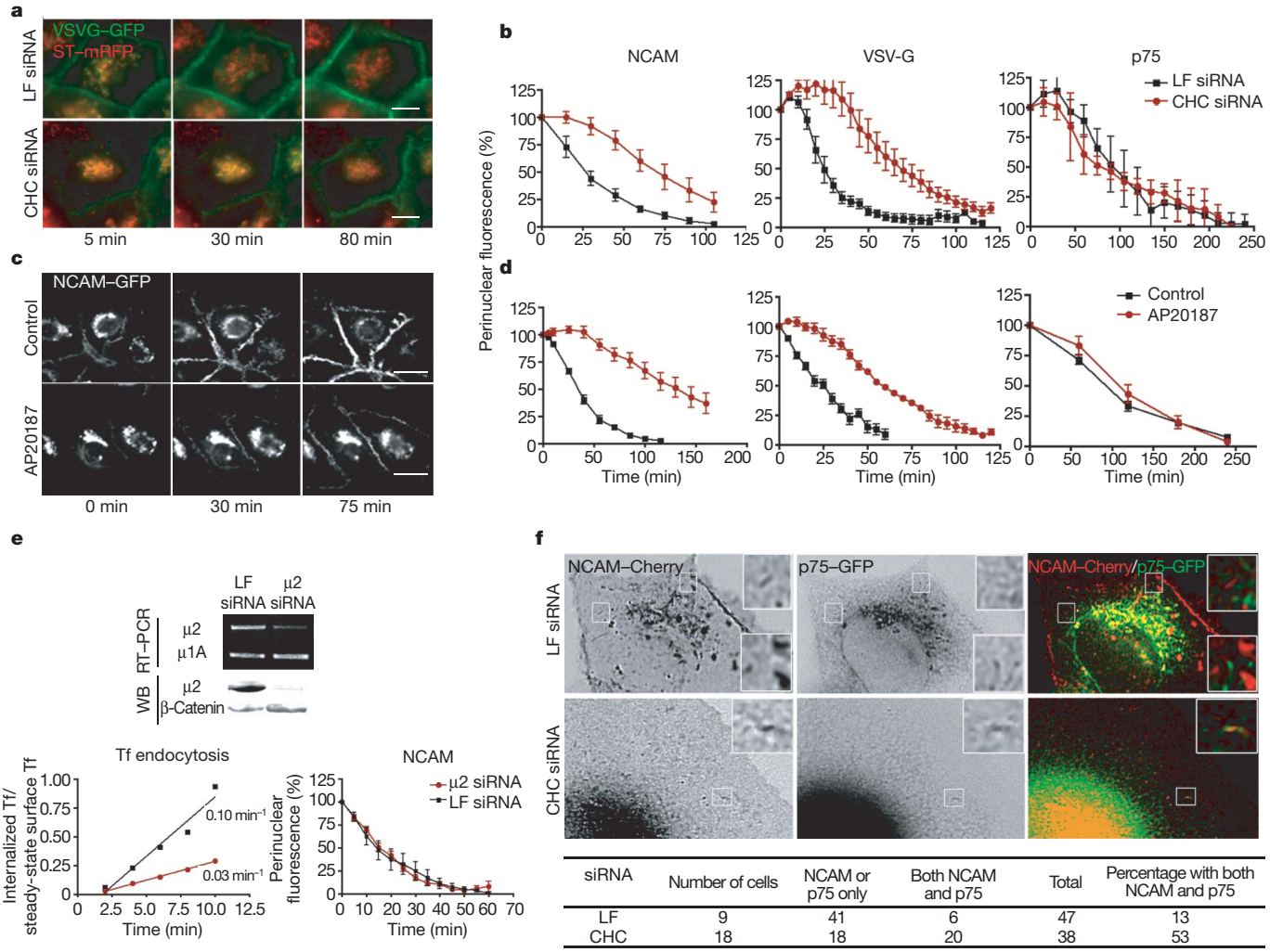
In yeast, knockdown of the CHC gene may be compensated for by the overexpression of a variety of genes[34], illustrating the marked ability of cells to develop compensatory mechanisms. We wished to test whether the apparent resistance of apical PM protein sorting to chronic clathrin depletion in MDCK cells might be due to selective compensation of clathrin function by mechanisms only available to apical proteins. To this end, we studied the effect of acutely blocking clathrin function on the exit of PM proteins from the TGN, by using a clathrin crosslinking approach previously shown to inhibit clathrin-mediated receptor endocytosis of TfR in non-polarized cells[15]. In brief, we generated an MDCK cell line in which CLC tagged with FK-506 binding protein (FKBP) and a haemagglutinin antigen (F-LC) replaced most of endogenous CLC (Supplementary Fig. 4a). Immunofluorescence showed that F-LC co-localized with CHC (Supplementary Fig. 4b). Addition of the membrane-permeant crosslinker of FKBP, AP20187, at 100 nM resulted in maximum clathrin clustering, as reported by a fluorescence titration assay (Supplementary Fig. 4c, d). Acute clathrin crosslinking did not cause a delay in the exit of apical protein p75 but caused a selective delay in the exit of basolateral proteins from the TGN, exactly as observed with chronic clathrin depletion by means of siRNA (Fig. 5c, d). Supplementary Movie 1 documents the delayed exit of NCAM–GFP in clathrin-crosslinked cells. Thus, these clathrin knockdown experiments show that clathrin is selectively required for the exit of newly synthesized basolateral PM proteins with tyrosine-dependent (VSVG[25]) and tyrosine-independent (NCAM[27]) sorting signals from the TGN, but not for the exit of an apical PM protein, p75 (refs 31, 33).

Basolateral proteins are mis-sorted into apical vesicles

The inhibitory effect of clathrin depletion on the exit of basolateral PM proteins from the Golgi apparatus could be explained by an indirect effect of the block of endocytosis and/or by the incorporation of the PM proteins into apical carrier vesicles that leave the Golgi complex with slower kinetics. To test the first hypothesis, we knocked down the PM clathrin adaptor AP2 in MDCK cells and conducted Golgi exit experiments similar to those shown in Fig. 5a–d. AP2 knockdown drastically inhibited the PM uptake of TfR but did not affect the exit of the basolateral marker NCAM–GFP from the TGN (Fig. 5e). These experiments showed that the inhibitory effect of clathrin depletion at the Golgi is not a result of inhibited endocytosis at the PM. To test the second hypothesis, we performed high-resolution live imaging experiments. These experiments showed that clathrin depletion promoted the incorporation of NCAM–Cherry into apical carrier vesicles containing the apical marker p75–GFP (53% of these vesicles showed the presence of NCAM–Cherry, in contrast with 13% in control cells) (Fig. 5f). Because p75–GFP leaves the Golgi complex with slower kinetics than that of NCAM (see Figs 1 and 2), our experiments show that clathrin depletion causes mis-sorting of basolateral proteins into a slower Golgi exit route normally used by apical proteins.

Clathrin is required for basolateral protein sorting

We have demonstrated a broad and selective requirement for clathrin in the biosynthetic sorting of basolateral PM proteins in epithelial cells. Accurate biosynthetic sorting is probably the main determinant of the polarity of basolateral PM proteins that are endocytosed slowly



siRNA	Number of cells	NCAM or p75 only	Both NCAM and p75	Total	Percentage with both NCAM and p75
LF	9	41	6	47	13
CHC	18	18	20	38	53

Figure 5 | Clathrin suppression disrupts exit and sorting of basolateral proteins at the TGN. a, Clathrin depletion by siRNA inhibits the exit of VSVG–GFP from the TGN. Experiments were performed in subconfluent MDCK cells expressing ST–RFP after a shift from 20 °C to 32 °C. **b**, Quantification of the exit of NCAM–GFP, VSVG–GFP and p75–GFP in control and clathrin-depleted cells. Perinuclear GFP was quantified as a fraction of the total cell GFP. Means ± s.e.m. are shown ($n > 10$; two experiments per condition). Scale bar, 10 μm. **c**, Clathrin crosslinking in MDCK cells expressing clathrin light chain linked to FKBP and haemagglutinin (HA). NCAM–GFP accumulated in the TGN after a 20 °C block moved to PM in control cells, but was blocked at the Golgi in cells treated with crosslinker AP20187 (see also Supplementary Movie 1). **d**, The exit of NCAM–GFP, VSVG–GFP and p75–GFP from the TGN was quantified from live recordings of MDCK cells expressing HA–FKBP-CLC, in the presence or absence of crosslinker. Data are expressed as means ± s.e.m. ($n > 10$ with at least two experiments per condition). **e**, Inhibition of endocytosis by AP2 knockdown does not delay the exit of NCAM–GFP from the TGN. Treatment of MDCK cells with μ2 siRNA reduced the Tf internalization rate by about 70% compared with control siRNA ($0.047 \pm 0.015\ min^{-1}$ versus $0.137 \pm 0.019\ min^{-1}$, $n = 4$) but did not affect NCAM–GFP exit from the TGN, quantified as in **b** and **d**. Means ± s.e.m. are shown. WB, western blot. **f**, Clathrin depletion promotes incorporation of basolateral protein into apical carriers. Cells were co-microinjected with cDNA encoding NCAM–Cherry (basolateral cargo) and p75–GFP (apical cargo). Post-Golgi carriers for p75–GFP, identified by their characteristic speed, contained NCAM–Cherry more often in clathrin-depleted cells than in control cells (represented in the boxed areas). See also Supplementary Movies 2 and 3.

relative to their recycling rates. Fast endocytic and recycling basolateral receptors, such as TfR and LDLR, also require clathrin in their basolateral recycling routes (Fig. 4a), in agreement with their dependence on AP1B for basolateral recycling[10,11]. These proteins are depolarized to a larger extent by clathrin depletion (Fig. 3), probably as a result of their dependence on this protein in both biosynthetic and recycling routes. The clathrin requirement of TfR in its biosynthetic route (Fig. 4b) is particularly interesting, because this receptor develops an AP1B-independent biosynthetic route in MDCK cells after two days in confluent culture[11]. Our observation predicts that a clathrin adaptor different from AP1B participates in the biosynthetic sorting of this receptor.

Clathrin-coated vesicles with specificity for different cargoes are produced at the plasma membrane, Golgi apparatus and endosomes through a cooperative process that requires clathrin and about 20 adaptors[35–39]. Besides AP1B, only one other AP-family adaptor, AP4, has been postulated to participate in basolateral PM protein delivery[40], but this adaptor does not seem to interact with clathrin[35–37]. Our finding that basolateral proteins that do not depend on AP1B for their basolateral polarity, for example CD147 and E-cadherin[11,28], require clathrin for their basolateral localization (Fig. 3a, b), also predicts the participation of other clathrin adaptors in basolateral polarity.

METHODS SUMMARY

Cell culture. MDCK cells were maintained in DMEM medium (Cellgro) containing 10% fetal bovine serum (Gemini) at 37 °C under 5% CO_2. For biotin polarity assays and immunostaining of polarized MDCK cells, cells were plated on Transwell polycarbonate filter supports with a pore size of 0.4 μm (Costar), and biotin assays were performed as described[11]. For microinjection experiments, MDCK cells were grown on coverslips and microinjected with the indicated cDNAs, as described previously[31,33].

siRNA and western blot analysis. siRNA duplexes were purchased from Dharmacon. Two CHC siRNA duplexes were used: 5′-UAAUCCAAUUCGAAGACCAAU-3′ (ref. 19) and 5′-GUAUGAUGCUGCUAAACUA-3′, designed with a Dharmacon program. Both oligonucleotides matched the predicted canine genome sequence of CHC. The μ2 siRNA duplex was 5′-GUGGAUGCCUUUCGGGUCA-3′. Sequences were subjected to BLAST searches to ensure that only the desired mRNA was targeted. Luciferase-specific siRNA (Dharmacon) was used as control. siRNAs were transfected into MDCK cells by nucleofection by Amaxa electroporation, with 5 μg (375 pmol) of siRNA and 4×10^6 cells, in accordance with the manufacturer's instructions (Amaxa). For effective knockdown, we performed two sequential nucleofections at 72-h intervals. Cells were then plated at subconfluence or confluence, depending on the experimental design, and the experiments were performed three days after the second round of nucleofection. Cell lysates containing 20 μg of protein were analysed by 10% SDS–PAGE and immunoblotting to detect expression levels of CHC, CLC and μ2. Normalization was performed with the signals obtained in parallel with antibody against β-coat protein or anti-β-catenin antibody.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 November 2007; accepted 7 February 2008.

1. Mostov, K., Su, T. & ter Beest, M. Polarized epithelial membrane traffic: conservation and plasticity. *Nature Cell Biol.* **5,** 287–293 (2003).
2. Rodriguez-Boulan, E., Kreitzer, G. & Musch, A. Organization of vesicular trafficking in epithelia. *Nature Rev. Mol. Cell Biol.* **6,** 233–247 (2005).
3. Yeaman, C., Grindstaff, K. K. & Nelson, W. J. New perspectives on mechanisms involved in generating epithelial cell polarity. *Physiol. Rev.* **79,** 73–98 (1999).
4. Brewer, C. B. & Roth, M. G. A single amino acid change in the cytoplasmic domain alters the polarized delivery of influenza viral hemagglutin. *J. Cell Biol.* **114,** 413–421 (1991).
5. Matter, K., Hunziker, W. & Mellman, I. Basolateral sorting of LDL receptor in MDCK cells: the cytoplasmic domain contains two tyrosine-dependent targeting determinants. *Cell* **71,** 741–753 (1992).
6. Le Bivic, A. *et al.* An internal deletion in the cytoplasmic tail reverses the apical localization of human NGF receptor in transfected MDCK cells. *J. Cell Biol.* **115,** 607–618 (1991).
7. Hunziker, W. & Fumey, C. A di-leucine motif mediates endocytosis and basolateral sorting of macrophage IgG Fc receptors in MDCK cells. *EMBO J.* **13,** 2963–2967 (1994).
8. Ohno, H. *et al.* Mu1B, a novel adaptor medium chain expressed in polarized epithelial cells. *FEBS Lett.* **449,** 215–220 (1999).
9. Folsch, H., Ohno, H., Bonifacino, J. S. & Mellman, I. A novel clathrin adaptor complex mediates basolateral targeting in polarized epithelial cells. *Cell* **99,** 189–198 (1999).
10. Gan, Y., McGraw, T. E. & Rodriguez-Boulan, E. The epithelial-specific adaptor AP1B mediates post-endocytic recycling to the basolateral membrane. *Nature Cell Biol.* **4,** 605–609 (2002).
11. Gravotta, D. *et al.* AP1B sorts basolateral proteins in recycling and biosynthetic routes of MDCK cells. *Proc. Natl Acad. Sci. USA* **104,** 1564–1569 (2007).
12. Cancino, J. *et al.* Antibody to AP1B adaptor blocks biosynthetic and recycling routes of basolateral proteins at recycling endosomes. *Mol. Biol. Cell* **18,** 4872–4884 (2007).
13. Folsch, H., Pypaert, M., Schu, P. & Mellman, I. Distribution and function of AP-1 clathrin adaptor complexes in polarized epithelial cells. *J. Cell Biol.* **152,** 595–606 (2001).
14. Rothman, J. E. & Fine, R. E. Coated vesicles transport newly synthesized membrane glycoproteins from endoplasmic reticulum to plasma membrane in two successive stages. *Proc. Natl Acad. Sci. USA* **77,** 780–784 (1980).
15. Moskowitz, H. S., Heuser, J., McGraw, T. E. & Ryan, T. A. Targeted chemical disruption of clathrin function in living cells. *Mol. Biol. Cell* **14,** 4437–4447 (2003).
16. Pagano, A., Crottet, P., Prescianotto-Baschong, C. & Spiess, M. *In vitro* formation of recycling vesicles from endosomes requires adaptor protein-1/clathrin and is regulated by rab4 and the connector rabaptin-5. *Mol. Biol. Cell* **15,** 4990–5000 (2004).
17. Li, J. *et al.* An ACAP1-containing clathrin coat complex for endocytic recycling. *J. Cell Biol.* **178,** 453–464 (2007).
18. Hinrichsen, L., Harborth, J., Andrees, L., Weber, K. & Ungewickell, E. J. Effect of clathrin heavy chain- and α-adaptin-specific small inhibitory RNAs on endocytic accessory proteins and receptor trafficking in HeLa cells. *J. Biol. Chem.* **278,** 45160–45170 (2003).
19. Motley, A., Bright, N. A., Seaman, M. N. & Robinson, M. S. Clathrin-mediated endocytosis in AP-2-depleted cells. *J. Cell Biol.* **162,** 909–918 (2003).
20. Royle, S. J., Bright, N. A. & Lagnado, L. Clathrin is required for the function of the mitotic spindle. *Nature* **434,** 1152–1157 (2005).
21. Iversen, T. G., Skretting, G., van Deurs, B. & Sandvig, K. Clathrin-coated pits with long, dynamin-wrapped necks upon expression of a clathrin antisense RNA. *Proc. Natl Acad. Sci. USA* **100,** 5175–5180 (2003).
22. Cereijido, M., Robbins, E. S., Dolan, W. J., Rotunno, C. A. & Sabatini, D. D. Polarized monolayers formed by epithelial cells on a permeable and translucent support. *J. Cell Biol.* **77,** 853–880 (1978).
23. Nelson, W. Adaptation of core mechanisms to generate cell polarity. *Nature* **422,** 766–774 (2003).
24. Odorizzi, G. & Trowbridge, I. S. Structural requirements for basolateral sorting of the human transferrin receptor in the biosynthetic and endocytic pathways of Madin–Darby canine kidney cells. *J. Cell Biol.* **137,** 1255–1264 (1997).
25. Thomas, D. C., Brewer, C. B. & Roth, M. G. Vesicular stomatitis virus glycoprotein contains a dominant cytoplasmic basolateral sorting signal critically dependent upon a tyrosine. *J. Biol. Chem.* **268,** 3313–3320 (1993).
26. Miranda, K. C. *et al.* A dileucine motif targets E-cadherin to the basolateral cell surface in Madin–Darby canine kidney and LLC-PK1 epithelial cells. *J. Biol. Chem.* **276,** 22565–22572 (2001).
27. Le Gall, A. H., Powell, S. K., Yeaman, C. A. & Rodriguez-Boulan, E. The neural cell adhesion molecule expresses a tyrosine-independent basolateral sorting signal. *J. Biol. Chem.* **272,** 4559–4567 (1997).
28. Deora, A. A. *et al.* The basolateral targeting signal of CD147 (EMMPRIN) consists of a single leucine and is not recognized by retinal pigment epithelium. *Mol. Biol. Cell* **15,** 4148–4165 (2004).
29. Nelson, W. J. & Veshnock, P. J. Ankyrin binding to (Na^+, K^+)ATPase and implications for the organization of membrane domains in polarized cells. *Nature* **328,** 533–536 (1987).
30. Herzlinger, D. A. & Ojakian, G. K. Studies on the development and maintenance of epithelial cell surface polarity with monoclonal antibodies. *J. Cell Biol.* **98,** 1777–1787 (1984).
31. Kreitzer, G., Marmorstein, A., Okamoto, P., Vallee, R. & Rodriguez-Boulan, E. Kinesin and dynamin are required for post-Golgi transport of a plasma-membrane protein. *Nature Cell Biol.* **2,** 125–127 (2000).
32. Raiborg, C., Wesche, J., Malerod, L. & Stenmark, H. Flat clathrin coats on endosomes mediate degradative protein sorting by scaffolding Hrs in dynamic microdomains. *J. Cell Sci.* **119,** 2414–2424 (2006).
33. Kreitzer, G. *et al.* Three-dimensional analysis of post-Golgi carrier exocytosis in epithelial cells. *Nature Cell Biol.* **5,** 126–136 (2003).
34. Lemmon, S. K. & Jones, E. W. Clathrin requirement for normal growth of yeast. *Science* **238,** 504–509 (1987).
35. Bonifacino, J. S. & Traub, L. M. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72,** 395–447 (2003).
36. Owen, D. J., Collins, B. M. & Evans, P. R. Adaptors for clathrin coats: structure and function. *Annu. Rev. Cell Dev. Biol.* **20,** 153–191 (2004).
37. Kirchhausen, T. Clathrin. *Annu. Rev. Biochem.* **69,** 699–727 (2000).
38. Conner, S. D. & Schmid, S. L. Regulated portals of entry into the cell. *Nature* **422,** 37–44 (2003).
39. Pearse, B. M., Smith, C. J. & Owen, D. J. Clathrin coat construction in endocytosis. *Curr. Opin. Struct. Biol.* **10,** 220–228 (2000).
40. Simmen, T., Honing, S., Icking, A., Tikkanen, R. & Hunziker, W. AP-4 binds basolateral signals and participates in basolateral sorting in epithelial MDCK cells. *Nature Cell Biol.* **4,** 154–159 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Müsch, T. McGraw and T. Ryan for comments on the manuscript. This work received support from NIH grants to E.R.-B., from the Dyson Foundation and from the Research to Prevent Blindness Foundation. E.P. was supported by a fellowship from Association pour la Recherche sur le Cancer during the initial stages of this work. R.S. was supported by a training predoctoral grant from the NIH.

Author Contributions S.D. designed research; S.D., E.P., D.G., R.S., S.B.S. and A.D. performed research; S.D. and E.R.-B. analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.R.-B. (boulan@med.cornell.edu).

METHODS

cDNA constructs and adenoviruses. Plasmids encoding NCAM–GFP, p75–GFP and ST–RFP were described previously[41,42]. NCAM–Cherry plasmid was generated by subcloning NCAM in pCherry vector into the *Sal*I and *Kpn*I site. pCherry vector was generated by excising GFP from pEGFP-N1 vector with *Bam*H1 and *Not*I enzymes and replacing it with Cherry. This vector was a gift from A. Müsch. Adenoviruses encoding the human TfR and VSVGtsO45–GFP were provided by I. Mellman and K. Simons, respectively. An MDCK cell line permanently expressing hPar6a–GFP was obtained from A. Müsch.

Antibodies. CHC was detected with the monoclonal antibody X22 (Affinity BioReagents) for immunoprecipitation and immunofluorescence experiments, and the monoclonal antibody TD1 (Covance) for immunofluorescence and western blot experiments. Clathrin light chains a and b were detected with the monoclonal antibody CON.1 (Covance). Mouse monoclonal antibody against haemagglutinin (HA) epitope and rabbit polyclonal antibody against giantin were purchased from Covance. Monoclonal antibody against ZO1 was purchased from Chemicon. Monoclonal anti-chicken NCAM antibody (clone 5e) was obtained from Developmental Studies Hybridoma Bank. Monoclonal antibody against the ectodomain of p75 (hybridoma ME 20.4) was provided by M. Chao. Monoclonal antibody against the ectodomain of VSVG (hybridoma 5FαG) was provided by J. Lewis. Monoclonal antibody against E-cadherin was purchased from BD Transduction Laboratories. Monoclonal antibody against Na^+, K^+-ATPase β subunit was purchased from Upstate.

Mouse anti-(human TfR) antibodies were purchased from Zymed (now Invitrogen) and from Chemicon. Goat antibody against human cathepsin D was purchased from R&D Systems. Antibody against β-coat protein was obtained from J. Rothman. Monoclonal antibody against the ectodomain of rat CD147 (RET-PE2 hybridoma) was provided by C. Barnstable. Rabbit anti-GFP antibody and chicken polyclonal anti-μ2 antibody were purchased from Abcam and rabbit anti-β-catenin antibody was purchased from Sigma.

Cathepsin D maturation assay. Cells grown in six-well plates were incubated with methionine-free DMEM for 30 min before labelling. After washes, the cells were pulse labelled with [^{35}S]methionine/cysteine in methionine/cysteine-free DMEM and then chased with normal DMEM supplemented with 5 mM methionine and 1 mM cysteine for different periods. After the chase, the cells were washed with ice-cold PBS and then lysed. Cathepsin D was immunoprecipitated with a goat anti-(human cathepsin D) antibody. Immunoprecipitates were analysed by SDS–PAGE followed by autoradiography. The band intensities were quantified with a phosphorImager (Molecular Dynamics).

Unidirectional flux of inulin. Inulin permeability was measured by incubating cell monolayers in the presence of 0.5 μCi of [^{3}H]inulin in the apical chamber. At various times, 50-μl aliquots of basal medium were withdrawn and radioactivity was measured with a Wallac Beta Counter (Perkin Elmer). Flux of [^{3}H]inulin into the basal well was calculated as the percentage of total radioactivity administered into the apical well.

Immunofluorescence and confocal microscopy. Three-day-old filter-grown monolayers of MDCK cells were fixed for 15–20 min in 4% paraformaldehyde in PBS containing Ca^{2+} and Mg^{2+}, and processed for indirect immunofluorescence by using primary and secondary antibodies tagged with Alexa568 or Alexa-488. Monolayers were stained intact (NCAM, VSVG, p75, CD147) or after permeabilization with 0.075% saponin (E-cadherin, clathrin X22, ZO1). Methanol fixation was used to stain the Na^+, K^+-ATPase β subunit and clathrin (with TD1 antibody). Confocal images were obtained with a laser-scanning confocal microscope TLS SP2 (Leica) with a 63× oil-immersion objective. Images were acquired with laser excitation bands at 488 and 543 nm.

Biotinylation assay to measure steady-state localization of PM proteins. To determine the steady-state surface polarity of PM proteins, MDCK-cell monolayers, confluent on Transwell chambers for three days, were biotinylated with sulpho-NHS-SS-biotin twice for 15 min at 4 °C, either apically or basolaterally. After biotinylation, the filters were excised and the cells were solubilized in lysis buffer (150 mM NaCl, 20 mM Tris pH 8.0, 5 mM EDTA, 1% Triton X-100, 0.2% BSA, and protease inhibitors). Biotinylated proteins were then incubated with streptavidin-agarose beads (Pierce) for 1 h. After washing of the beads, the biotinylated proteins were eluted in SDS-containing sample buffer (Laemmli) and analysed by SDS–PAGE and western blotting with specific antibodies.

Biotin targeting assays. Quantification of polarized delivery of TfR to apical and basolateral surfaces of MDCK cells was performed as described previously[11]. MDCK cells confluent on filters for two days were infected with adenovirus encoding human TfR for 28–32 h and were subsequently pulse-labelled for 13 min with 0.5 mCi ml^{-1} [^{35}S]methionine/cysteine. The excess radioactivity was quickly rinsed off with serum-free DMEM supplemented with 3 mM unlabelled methionine and cysteine. Cells were then chased for the indicated times at 37 °C with serum-free DMEM supplemented with 50 mg ml^{-1} biotinylated human transferrin (B-Tf) added to either the apical or basolateral sides. At the end of the chase period, excess B-Tf was removed by three consecutive rinses with Hanks balanced salt solution (HBSS) supplemented with 0.5% BSA. The cells were then lysed for 30 min at 4 °C in TfR lysis buffer (TfR-LB/5.5) containing 150 mM NaCl, 30 mM KCl, 2.5 mM EDTA, 1.5% Triton-X100 and 1% albumin in 40 mM sodium acetate buffer pH 5.5 and supplemented with a protease inhibitor cocktail that included 1 mM phenylmethylsulphonyl fluoride, 15 μg ml^{-1} leupeptin/pepstatin/antipain and 75 μg ml^{-1} benzamidine-ClH. The B-Tf–TfR complex was then retrieved from the lysate with avidin-Sepharose, boiled in Laemmli sample buffer and resolved on 9% SDS–PAGE. The roughly 90-kDa band corresponding to TfR was detected by phosphorImager and quantified.

To examine the polarized delivery of VSVG to the cell surface, MDCK cells grown for three days on Transwell chambers were infected with adenovirus encoding VSVGtsO45–GFP. After 24 h, the monolayers were pulse-labelled for 15 min with 0.5 mCi ml^{-1} [^{35}S]methionine/cysteine. At various chase times, filters were chilled to 4 °C and biotinylated from the apical or basolateral sides; the lysates were subjected to immunoprecipitation with specific antibodies against VSVG protein. Samples were subsequently re-precipitated with streptavidin-agarose, and the biotin-labelled immunoprecipitates were separated by SDS–PAGE, revealed with a phosphorImager and quantified. For CD147 we used a cell line expressing CD147–GFP, and immunoprecipitation was performed with an anti-GFP antibody.

^{125}I-Tf internalization assay. One day before the experiment, MDCK cells were infected with adenovirus encoding human TfR. MDCK cells were incubated for 2, 4, 6, 8 or 10 min with 5 μg ml^{-1} ^{125}I-labelled human Tf at 37 °C. After incubation, cells were placed on ice, washed twice with cold neutral pH buffer (Ca/Mg HBSS buffer; Gemini) and incubated for 5 min at 4 °C in pH 2 buffer (0.5 M NaCl, 0.5 M acetic acid pH 2) to remove surface Tf. After a quick wash with neutral pH buffer, the cells were solubilized and radioactivity was measured. A parallel plate was incubated for 2 h in neutral pH buffer with 5 μg ml^{-1} ^{125}I-labelled human Tf at 4 °C to determine the total surface amount of receptors. The cells were washed six times with neutral pH buffer and solubilized (1% Triton X-100, 0.1 M NaOH), and radioactivity was measured with a Wallac Gamma Counter (Perkin Elmer). Non-specific binding was measured in the presence of a 200-fold excess of unlabelled human Tf and subtracted. The slope of a plot of the ratio of internal human Tf to surface human Tf (4 °C binding) against time was used to measure the internalization time constant.

Recycling of TfR and LDLR. Recycling of TfR was measured with ^{125}I-labelled human Tf (^{125}I-Tf) as a ligand in MDCK cells transduced with an adenovirus encoding human TfR, as described previously[11]. For these experiments, MDCK cells confluent on Transwell filters for four days were exposed to ^{125}I-Tf from the basolateral side for 2 h, washed at low pH to remove surface ^{125}I-Tf and incubated for different durations at 37 °C in the presence of unlabelled Tf and desferroxamine to follow the release of ^{125}I-Tf into basal medium (basolateral recycling) and apical medium (transcytosis). Recycling of LDLR was studied in MDCK cells permanently expressing BBS-LDLR (α-bungarotoxin (BTX) binding site fused to LDLR), a fusion protein in which part of the luminal domain of human LDLR was replaced by a peptide-binding sequence for bungarotoxin (BBS), thus allowing the use of ^{125}I-BTX as a high-affinity ligand. Cells confluent on filters for four days were incubated for 3 h with 5 μCi ml^{-1} (0.29 μg ml^{-1}) [^{125}I]Tyr^{54}-BTX (PerkinElmer) at 37 °C to achieve steady-state occupancy of the tagged LDL receptors. The cells were washed twice with ice-cold Hanks buffer with calcium and magnesium. The surface pool of ^{125}I-BTX bound to the BBS-LDLR was cleaved off with 25 μg ml^{-1} trypsin in Hanks buffer at 4 °C for 30 min. After being washed, the filters were placed at 37 °C in the presence of 25 μg ml^{-1} trypsin added apically and basolaterally. The medium was collected at 7, 15, 27, 42, 60, 80 or 100 min from apical and basolateral chambers, and radioactivity was measured with a Wallac Gamma Counter. Each time point was performed in triplicate.

Microinjection, live imaging of exit from the TGN and quantification of fluorescence. Expression and accumulation in the TGN of p75–GFP and NCAM–GFP were performed as described previously[31]. MDCK cells grown on glass coverslips were microinjected in their nuclei with 1–20 μg ml^{-1} cDNA encoding the GFP-tagged proteins, diluted in HKCl injection buffer (10 mM HEPES, 140 mM KCl pH 7.4) and incubated at 37 °C for 1–2 h and at 20 °C with 100 μg ml^{-1} cycloheximide for accumulation in the TGN. After release of the temperature block by incubation at 32 °C, cells were imaged as described[31]. Expression of VSVG–GFP was obtained by transduction with an adenovirus encoding VSVGtsO45–GFP 24 h before the experiment. The cells were incubated overnight at 39 °C for synthesis and accumulation of the protein at the endoplasmic reticulum and incubated for 2 h at 20 °C in the presence of 100 μg ml^{-1} cycloheximide. The Golgi block was released by transfer to 32 °C and perinuclear and total GFP fluorescence were quantified as described[31], to assess the exit of

NCAM–GFP, p75–GFP or VSVG–GFP from the TGN. Imaging was conducted with a temperature-controlled stage and a TE300 inverted microscope (Nikon) with a 40× (1.04 numerical aperture) objective. Time-lapse movies were acquired with an ORCA II ER camera (Hamamatsu Photonics).

For dual-colour movies, cells were co-microinjected with p75–GFP ($1\,\mu g\,ml^{-1}$) and NCAM–Cherry ($20\,\mu g\,ml^{-1}$). The co-localization analysis of NCAM and p75 exocytic cargoes was performed manually by tracking single-positive and double-positive structures in the acquired time sequences with the use of Metamorph software (Molecular Devices).

Clathrin crosslinking. The fusion protein construct HA–FKBP-LC was prepared from a plasmid encoding GFP–FKBP-LC, obtained from T. Ryan[15]. The FKBP portion of this construct is modified to specifically bind the crosslinker AP20187 (Ariad Pharmaceuticals; http://www.ariad.com/regulationkits). To fit our experimental design, the GFP tag was replaced by an HA tag. The sequence encoding GFP was removed with *Nhe*I and *Bgl*II and the sequence encoding HA was added by ligating the following insert phosphorylated at the 5′ end:

5′-CTAGCTCGCCACCATGTACCCATATGACGTCCCCGACTA-CGCGA-3′
3′-GAGCGGTGGTACATGGGTATACTGCAGGGGCTGATGCG-CTCTAG-5′

The resulting plasmid, pHA–FKBP-LC, was stably transfected in MDCK cells with the use of the Amaxa nucleofection technique (Amaxa). MDCK clones were selected after 1–2 weeks of culture in the presence of G418 and assayed for expression of HA–FKBP-LC by immunofluorescence and western blotting with anti-HA and anti-CLC antibodies.

The ideal concentration of crosslinker (AP20187), added for 30–60 min at 37 °C, to obtain maximal clustering of perinuclear clathrin, was titrated using a Metamorph-based assay that involved measuring the intensity of a thresholded region in each cell, with the threshold set at 50% of the total intensity for one condition and then applied to other conditions.

Statistical analysis. Statistically significant differences between samples were determined by a paired two-tailed *t*-test analysis using Prism.

41. Deora, A. A., Diaz, F., Schreiner, R. & Rodriguez-Boulan, E. Efficient electroporation of DNA and protein into confluent and differentiated epithelial cells in culture. *Traffic* **8**, 1304–1312 (2007).
42. Musch, A., Cohen, D., Kreitzer, G. & Rodriguez-Boulan, E. cdc42 regulates the exit of apical and basolateral proteins from the trans-Golgi network. *EMBO J.* **20**, 2171–2179 (2001).

LETTERS

Thermodynamic control by frequent quantum measurements

Noam Erez[1], Goren Gordon[1], Mathias Nest[2] & Gershon Kurizki[1]

Heat flow between a large thermal 'bath' and a smaller system brings them progressively closer to thermal equilibrium while increasing their entropy[1]. Fluctuations involving a small fraction of a statistical ensemble of systems interacting with the bath result in deviations from this trend. In this respect, quantum and classical thermodynamics are in agreement[1–5]. Here we predict a different trend in a purely quantum mechanical setting: disturbances of thermal equilibrium between two-level systems (TLSs) and a bath[6], caused by frequent, brief quantum non-demolition[7-10] measurements of the TLS energy states. By making the measurements increasingly frequent, we encounter first the anti-Zeno regime and then the Zeno regime (namely where the TLSs' relaxation respectively speeds up and slows down[11-15]). The corresponding entropy and temperature of both the system and the bath are then found to either decrease or increase depending only on the rate of observation, contrary to the standard thermodynamical rules that hold for memory-less (Markov) baths[2,5]. From a practical viewpoint, these anomalies may offer the possibility of very fast control of heat and entropy in quantum systems, allowing cooling and state purification over an interval much shorter than the time needed for thermal equilibration or for a feedback control loop.

To understand the origins of the predicted anomalies, consider a thermal bath in equilibrium with an ensemble of quantum systems. The energy of the quantum systems is briefly measured, with the following effects. Classically, the equilibrium state of the systems may remain intact, since measurements can be chosen to be non-intrusive, that is, to involve no energy exchange, and merely to provide 'snapshots' of the overall system. Likewise, quantum mechanically, nearly ideal (projective) measurements involve no energy cost when performed by macroscopic detectors on isolated systems[8]. However, finite-time coupling, followed by abrupt decoupling, of two quantum ensembles, which may be viewed as the detection of one ensemble by the other, may cause an increase in their mean total energy[16]. Controlled perturbations of open quantum systems may also cause their properties to change anomalously[17,18].

Here we are concerned with the time-resolved evolution of two-level systems (TLSs) that are initially in thermal equilibrium with a much larger bath. We investigate how the temperature and entropy of these systems evolve when the systems are probed by brief, repeated measurements. Our analysis shows that the temperature and entropy display a universal dependence on the time interval between measurements, provided that they are frequent enough to cause deviations from energy conservation in the system–bath exchange[19], consistent with time–energy uncertainty.

Quantitatively, we consider a TLS with energy separation $\hbar\omega_a$ ('a' denoting 'atom') that is weakly coupled to a thermal bath of harmonic oscillators, characterized by a bath response or correlation (memory) time of $t_c \gg 1/\omega_a$, which typically marks the onset of equilibrium. After equilibrium has been reached, we perform $k = 1, \ldots, K$ quantum non-demolition measurements[7–10] of the TLS energy states at times separated by $\Delta t_k = t_{k+1} - t_k$. Each measurement has a brief duration of $\tau_k \ll 1/\omega_a$. Our aim is to explore the evolution as a function of the time separation $\Delta t_k \leq 1/\omega_a \ll t_c$ between consecutive measurements in the uncharted non-Markov domain.

We show that this process entails three universal anomalies. The first of these is that the quantum mechanical non-commutativity of the system–detector and system–bath interactions causes the system to heat up immediately after the measurement, at the expense of the detector–system coupling but not at the expense of the bath. This heating up occurs only for very low values of Δt_k compatible with the quantum Zeno effect (QZE)[11]. The second anomaly is that a transition of the TLS ensemble from heating to cooling may occur as we vary the interval between consecutive measurements from $\Delta t_k \ll 1/\omega_a$ to $\Delta t_k \approx 1/\omega_a \ll t_c$. This marks the transition from values of Δt_k compatible with the QZE to values compatible with the anti-Zeno effect (AZE)[12–15]. We note that the cooling may occur even if the bath is initially hotter. The third anomaly is that, correspondingly, oscillations of the entropy relative to that of the equilibrium state take place, contrary to the markovian notion of the second law of thermodynamics[2,5].

This generic scenario is governed by the following total hamiltonian of the system that interacts with the bath and is intermittently perturbed by the coupling of the system to the detector (measuring apparatus):

$$H(t) = H_{tot} + H_{SD}(t)$$

$$H_{tot} = H_S + H_B + H_{SB}$$

Here H_S is the hamiltonian of the TLS, with ground and excited states $|g\rangle$ and $|e\rangle$, respectively; H_B is that of the thermal bath composed of harmonic oscillators with energies $\hbar\omega_\lambda$ (λ denoting the mode of the bath); $H_{SB} = SB$ is the system–bath interaction hamiltonian[19] (the spin-boson interaction), which is a product of the system-dipole (or spin-flip) operator S and the operator B describing the bath excitations and de-excitations; and $H_{SD}(t)$ is the time-dependent measurement hamiltonian that couples the system to a detector consisting of energy-degenerate ancillae (for details, see Supplementary Information section A). It is essential that the coupling hamiltonians H_{SB} and H_{SD} do not invoke the rotating-wave approximation[19]; that is, we do not impose energy conservation between the system and the bath or the detector, on the timescales considered[15].

The near-equilibrium state ρ_{tot} before a measurement has three pertinent characteristics (see Supplementary Information section B). First, it displays system–bath entanglement with off-diagonal matrix elements $\langle e|\rho_{tot}|g\rangle \neq 0$. Second, the system is described by a diagonal reduced density matrix $\rho_S = Tr_B\rho_{tot}$ (Tr_B denoting trace over the bath) in the H_S eigenbasis. Third, the mean interaction energy $\langle H_{SB}\rangle$ is negative, assuming that ρ_{tot} weakly deviates from the ground

[1]Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Theoretische Chemie, Universitaet Potsdam, Potsdam 14476, Germany.

state of H_{tot}: $\langle H_{SB}\rangle = \langle H_{tot}\rangle - \langle H_S + H_B\rangle < 0$. This is because the correction to the ground-state energy of H_{tot} due to a weakly perturbing interaction H_{SB} is negative (to leading second order).

We next consider the disturbance of this equilibrium state by a nearly impulsive (projective) quantum measurement ($\tau \to 0$) of the TLS, in the basis $\{|g\rangle, |e\rangle\}$. Such measurements do not resolve the energies of the TLS states, due to time–energy uncertainty. However, they can discriminate between states of different symmetry, for example different angular momenta[19]. The measurement correlates the TLS energy eigenstates with mutually orthogonal states of an ancillary detector and has distinctly quantum mechanical consequences (see Supplementary Information section A): using the energy supplied by $H_{SD}(0 < t < \tau)$ (the system–detector coupling), but without changing $\langle H_D\rangle$, it eliminates the mean system–bath interaction energy; that is, it sets $\langle H_{SB}\rangle$, the pre-measurement value of which was negative, to zero. Thus, for $0 \le t \le \tau$:

$$\langle H_{SB}(0)\rangle < 0 \to \langle H_{SB}(\tau)\rangle = 0$$
$$\langle H_{SD}(t)\rangle = -\langle H_{SB}(t)\rangle$$

This energy transfer, resulting in a change in the entanglement between the system and the bath, triggers the quantum dynamics that redistributes their mean energy and entropy.

The information gained from such measurements may be used to sort system sub-ensembles according to their measured energy, to extract work or entropy change, in the spirit of Maxwell's demon[20]. Here, however, we let the entire TLS ensemble evolve independently of the measured result; that is, we trace out the detector states, a procedure known as non-selective, or unread, measurements[10].

After the measurement (because $H_{SD}(t \ge \tau) = 0$), time–energy uncertainty at $\Delta t \le 1/\omega_a$ results in the breakdown of the rotating-wave approximation; that is, $\langle H_S + H_B\rangle$ is not conserved as Δt grows. Only $\langle H_{tot}\rangle$ is conserved, by unitarity, until the next measurement. Hence, the post-measurement decrease of $\langle H_{SB}\rangle$ with Δt, signifying the restoration of equilibrium, $\langle H_{SB}(\tau)\rangle = 0 \to \langle H_{SB}(\tau + \Delta t)\rangle < 0$, is at the expense of an increase in $\langle H_S + H_B\rangle = \langle H_{tot}\rangle - \langle H_{SB}\rangle$, that is, a heating of the system and the bath combined:

$$\left.\frac{d}{dt}(\langle H_S\rangle + \langle H_B\rangle)\right|_{\tau+\Delta t} > 0$$
$$\left.\frac{d}{dt}\langle H_{SB}\rangle\right|_{\tau+\Delta t} < 0 \qquad (1)$$

The post-measurement evolution of the system alone, described by $\rho_S = \mathrm{Tr}_B\rho_{tot}$, is not obvious. Its Taylor expansion:

$$\rho_S(\tau+\Delta t) \approx \rho_S(\tau) + \Delta t\dot{\rho}_S(\tau) + \frac{\Delta t^2}{2}\ddot{\rho}_S(\tau) + \cdots \qquad (2)$$

where dots denote differentiation with respect to time, holds at short evolution times $\Delta t \ll 1/\omega_a$. The zeroth-order term is unchanged by the measurement: $\rho_S(\tau) = \rho_S(t \le 0)$. The first derivative vanishes at $t = \tau(\Delta t = 0)$ owing to the definite parity of the bath density operator correlated with $|g\rangle$ or $|e\rangle$ (see Supplementary Information section B). This initial post-measurement vanishing, $\dot{\rho}_S(\tau) = 0$, is the QZE condition[11,13–15]. The time evolution of ρ_S is then governed by its second time derivative, $\ddot{\rho}_S(\tau)$, which can be shown (see Supplementary Information section B) to have the same sign as $\sigma_z = |e\rangle\langle e| - |g\rangle\langle g|$, the z Pauli matrix, which is the population difference operator of the TLS. Hence, the second derivative in equation (2) is positive shortly after the measurement, consistent with equations (1), if there is no initial population inversion of the system, that is, for non-negative temperature.

The evolution of ρ_S at longer times (in the regime of weak system–bath coupling) may be approximately described (as verified by our exact numerical simulations[21,22]; see Supplementary Information section C) by the second-order non-markovian master equation (Fig. 1 main panel). Because ρ_S is diagonal, the master equation describing it can be cast into the following population-rate equations[15] (we drop the subscript 'S' in what follows and set the measurement time to be $t = 0$):

$$\dot{\rho}_{ee}(t) = -\dot{\rho}_{gg}(t) = R_g(t)\rho_{gg} - R_e(t)\rho_{ee} \qquad (3)$$

$$R_e(t) = 2t\int_{-\infty}^{\infty} d\omega G_T(\omega)\mathrm{sinc}[(\omega-\omega_a)t]$$
$$R_g(t) = 2t\int_{-\infty}^{\infty} d\omega G_T(\omega)\mathrm{sinc}[(\omega+\omega_a)t] \qquad (4)$$
$$G_T(\omega) = (n_T(\omega)+1)G_0(\omega) + n_T(-\omega)G_0(-\omega)$$

Here $\mathrm{sinc}(x) = \sin(x)/x$ is the Fourier transform of the time interval after the measurement, $G_T(\omega)$ is the temperature-dependent coupling spectrum of the bath, $G_0(\omega)$ is the zero-temperature coupling spectrum with peak coupling strength γ at ω_0 and spectral width $\sim 1/t_c$, and $n_T(\omega) = \left(e^{\beta\omega} - 1\right)^{-1}$ is the inverse-temperature-dependent (that is, β-dependent) population of bath mode ω.

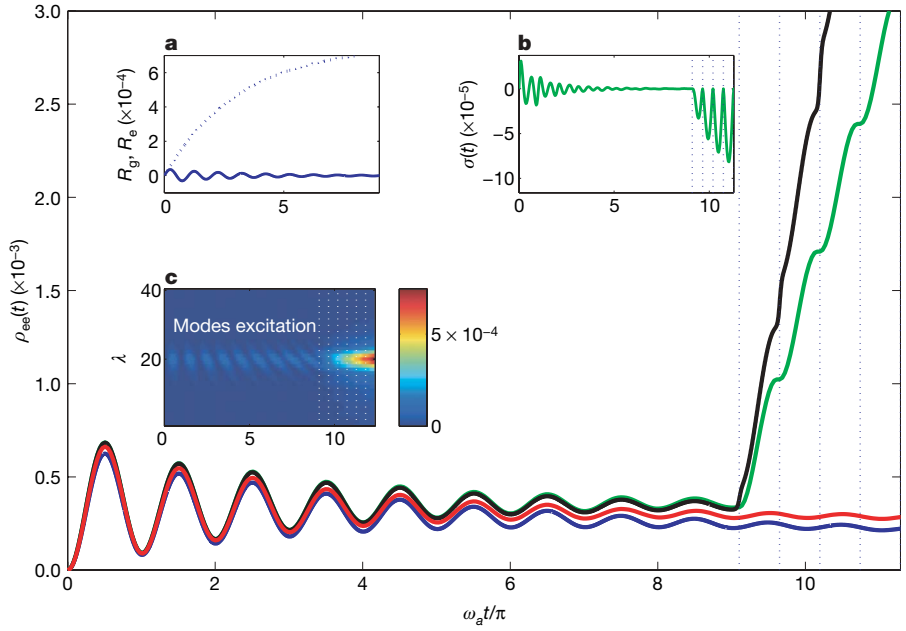


Figure 1 | System and bath evolution as a function of time. The main panel shows the excited-level population as a function of time for the initial zero-temperature product state of the system and bath. The population relaxes to quasi-equilibrium after a few oscillations. It is then subjected to a series of measurements (vertical dotted lines). Measurements of finite duration ($\tau_k = 0.11/\omega_a$) (black line) result in a larger heating up than do impulsive measurements (green line), but the dominant effect is the same for both. We observe agreement between the results of the second-order master equation (blue line), two-quanta exchange with a discrete bath (black and green lines), and the exact numerical solution for a discrete bath of 40 modes (red line). The horizontal axes of the inset panels all indicate $\omega_a t/\pi$, as in the main panel: **a**, relaxation rates R_g (solid line) and R_e (dotted line) as functions of time; **b**, negative of the rate of change of relative entropy, $\sigma(t)$; **c**, excitations of the 40 bath modes in the two-quanta model as functions of time. Parameters: $t_c = 10/\omega_a$, $\omega_0 = \omega_a$, $\gamma = 0.07\omega_a$.

The dynamics of equation (3) is determined by $R_e(t)$ and $R_g(t)$, which are respectively the relaxation rates of the excited and ground states. Their non-Markov time dependence yields three distinct regimes, universally dependent on the post-measurement times t of the spin-boson evolution.

First, at short times $t \ll 1/\omega_a \ll t_c$, the sinc functions in equations (4) are spectrally much broader than is G_T. The relaxation rates R_e and R_g are thus equal at any temperature, indicating the complete breakdown of the rotating-wave approximation discussed above: the $|g\rangle \rightarrow |e\rangle$ and $|e\rangle \rightarrow |g\rangle$ transitions do not require quantum absorption or emission by the bath, respectively. The rates R_e and R_g thus become linear in time, manifesting the QZE[13–15]:

$$R_e(t \ll t_c) \approx 2\dot{R}_0 t$$

$$R_g(t \ll t_c) \approx 2\dot{R}_0 t$$

$$\dot{R}_0 \equiv \int_{-\infty}^{\infty} d\omega G_T(\omega) = \langle B^2 \rangle$$

This implies that in the short-time regime we have the universal Zeno heating rate:

$$\frac{d}{dt}\left(\rho_{ee} - \rho_{gg}\right) \approx 4\dot{R}_0 t\left(\rho_{gg} - \rho_{ee}\right)$$

Second, at intermediate non-markovian times $t \approx 1/\omega_a$, when the sinc functions and G_T in equations (4) have comparable widths, the relaxation rates R_e and R_g exhibit several unusual phenomena that stem from time–energy uncertainty. The change in the overlap of the sinc and G_T functions with time results in damped aperiodic oscillations of $R_e(t)$ and $R_g(t)$, near the frequencies $\omega_0 - \omega_a$ and $\omega_0 + \omega_a$, respectively. We call this oscillatory time dependence, which is displayed in neither the QZE nor the converse AZE of accelerated relaxation[12–14], the oscillatory Zeno effect (OZE). Owing to the negativity of the sinc function between its consecutive maxima, we can have a negative relaxation rate, which is completely forbidden by the rotating-wave approximation. Because $\mathrm{sinc}[(\omega + \omega_a)t]$ is shifted much further from the peak of $G_T(\omega)$ than is $\mathrm{sinc}[(\omega - \omega_a)t]$, $R_g(t)$ is more likely to be negative than is $R_e(t)$ (Figs 1a, 2a). Hence, $\rho_{gg}(t)$ may grow at the expense of $\rho_{ee}(t)$ more than is allowed by the thermal-equilibrium detailed balance. This may cause transient cooling, as described below.

Third, at long times $t \gg t_c$, the relaxation rates attain their 'golden rule' (Markov) values[15]:

$$R_e(t \gg t_c) \approx 2\pi G_T(+\omega_a)$$

$$R_g(t \gg t_c) \approx 2\pi G_T(-\omega_a)$$

The populations thus approach those of an equilibrium Gibbs state the temperature of which is equal to that of the thermal bath (Fig. 1 main panel).

We now consider entropy dynamics. The entropy of ρ_S relative to its equilibrium state ρ_0 ('entropy distance'), and the negative of its rate of change, can always be defined[3,5]:

$$S(\rho_S(t)\|\rho_0) \equiv \mathrm{Tr}\{\rho_S(t)\ln\rho_S(t)\} - \mathrm{Tr}\{\rho_S(t)\ln\rho_0\}$$

$$\sigma(t) \equiv -\frac{d}{dt}S(\rho_S(t)\|\rho_0) \qquad (5)$$

In the markovian realm, σ thus defined is identified as the 'entropy production rate'[2,3,5], and $\sigma(t) \geq 0$ is a statement of the second law of thermodynamics in this realm. Because ρ_S is diagonal, it follows (see Supplementary Information section D) that $\sigma(t)$ is positive if and only if $d\left(\left|\rho_{ee}(t) - (\rho_0)_{ee}\right|\right)/dt \leq 0$, consistent with the interpretation of the relative entropy $S(\rho_S\|\rho_0)$ in equations (5) as the entropic distance from equilibrium. However, in the short-time regime, whenever the oscillatory $\rho_{ee}(t)$ drifts away from its initial or final equilibrium σ takes negative values (Fig. 1b) that are forbidden by the markovian statement of the second law.

To realistically model the repeated measurements, that is, give them finite duration, we assume a smooth temporal profile of the

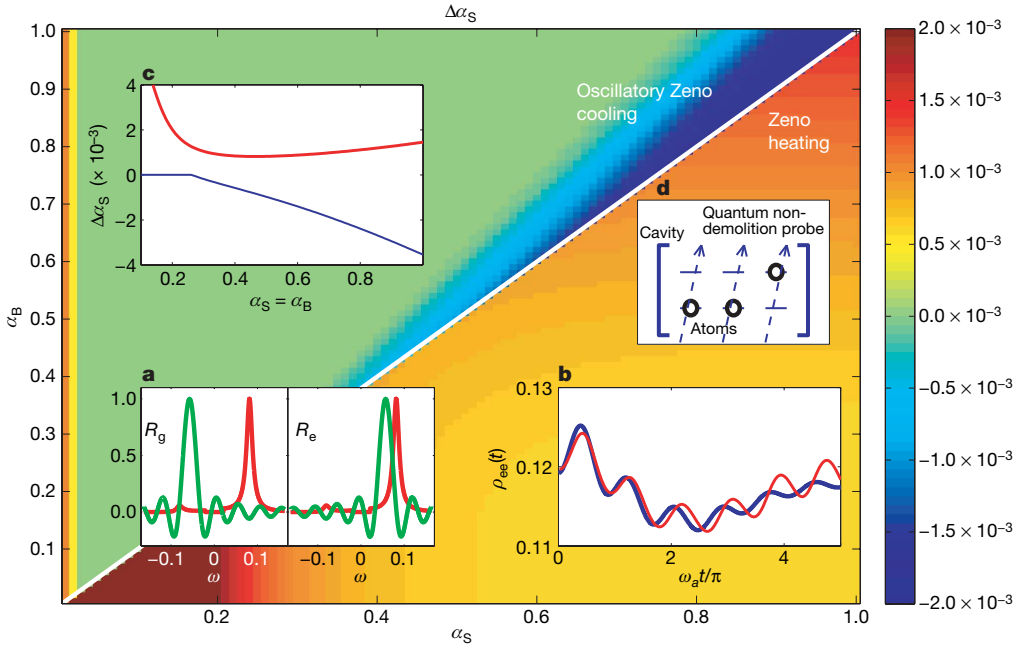


Figure 2 | Maximal heating and cooling of the system. The main panel shows the maximal heating (lower half) and cooling (upper half) of the system, for different system (horizontal axis) and bath (vertical axis) initial temperatures α_S and α_B ($\alpha_S \neq \alpha_B$), where $\alpha_S = \hbar\omega_a/\beta_S$, $\alpha_B = \hbar\omega_a/\beta_B$, and $\Delta\alpha_S = \max(\alpha_S(t) - \alpha_S(0))$ for heating and $\Delta\alpha_S = \min(\alpha_S(t) - \alpha_S(0))$ for cooling. (β denotes the inverse temperature.) **a**, R_g and R_e (see equations (4)), depicted as spectral overlaps of $G_T(\omega)$ (red) and $\mathrm{sinc}[(\omega + \omega_a)t]$ (for R_g) and $\mathrm{sinc}[(\omega - \omega_a)t]$ (for R_e) (green). **b**, Example of a system undergoing first Zeno heating and then oscillatory Zeno cooling, obtained from the second-order master equation (blue) and from the exact numerical solution for a discrete bath of 40 modes (red). **c**, Maximal Zeno heating (red) and subsequent maximal oscillatory Zeno cooling (blue), as functions of the common initial temperature of the system and the bath $\alpha_S = \alpha_B$. We note the presence of a critical temperature for oscillatory Zeno cooling. Parameters: $t_c = 10/\omega_a$, $\omega_0 = \omega_a/0.7$, $\gamma = 4.36\omega_a$. These effects can be strongly magnified by choosing other suitable parameters. **d**, Possible experimental setup.

coupling to the detector (see Supplementary Information section A). The kth measurement then occurs at time t_k and has a duration of τ_k. Figure 1(main panel) compares the population evolution using projective (impulsive) and finite-duration measurements with $\tau_k \approx 0.1/\omega_a$. Finite-duration measurements increase the Zeno heating in comparison with impulsive ones on account of the extra energy supplied by the coupling to the detector. However, the basic effect is seen (Fig. 1 main panel) to be the same and is governed by the time derivative of $\langle H_{SB} \rangle$ (see equations (1)). Counterintuitively, finite-duration measurements can increase the cooling, despite the extra energy supplied by the apparatus.

If we repeat this procedure often enough, the TLSs will increasingly either heat up or cool down, if we choose the time intervals Δt_k to coincide with either peaks or troughs of the ρ_{ee} oscillations, respectively. The minimal value of σ can also be progressively lowered with each measurement (Fig. 1b). Because consecutive measurements affect the bath and the system differently, they may acquire different excitations or de-excitations, which then become the initial conditions for subsequent QZE heating or OZE cooling. The results are shown in Fig. 2 for both different (main panel) and common (Fig. 2c) initial temperatures of the system and the bath. We note that the system may heat up on account solely of the QZE, although the bath is initially colder, or cool down on account solely of the OZE or AZE, although the bath is initially hotter (Fig. 2 main panel). The bath may also undergo changes in energy and entropy (Fig. 1c).

One experimental realization of these effects can involve atoms or molecules in a microwave cavity (Fig. 2d) with a coupling spectrum $G_T(\omega)$ (controllable using the cavity quality factor and temperature) centred at ω_0. Measurements can be effected on such a TLS ensemble with resonance frequency ω_a in the microwave domain, at time intervals $\Delta t_k \approx 1/(\omega_0 \pm \omega_a)$, using an optical quantum non-demolition probe[8] at a frequency ω_p such that $\omega_p \gg \omega_a$ and $\omega_p \gg \omega_0$. The probe pulses undergo a different Kerr-nonlinear phase shift $\Delta\phi_e$ or $\Delta\phi_g$ depending on the different symmetries (for example angular momenta) of $|e\rangle$ and $|g\rangle$. The relative abundance of $\Delta\phi_e$ and $\Delta\phi_g$ thus reflects the ratio $\rho_{ee}(t_k)/\rho_{gg}(t_k)$. Such quantum non-demolition probing may be performed with time duration much shorter than ω_a^{-1}, that is, $\omega_a \tau_k \ll 1$, without resolving the energies of $|e\rangle$ and $|g\rangle$. Ion traps or solid matrices may serve as phonon baths instead of photonic cavities.

Since non-selective measurements increase the von Neumann entropy of the detector ancillae, their entropic price precludes a 'perpetuum mobile' if closed-cycle operation is attempted. However, if our ancillae are laser pulses, they are only used once and we may progressively change the TLS ensemble thermodynamics by consecutive pulses, disregarding their entropic or energetic price. The practical advantage of the predicted anomalies is the possibility of very rapid control of cooling and entropy, which may be attained after several measurements at $t \geq \omega_a^{-1}$ and is only limited by the measurement rate. By contrast, conventional cooling requires much longer times, $t \gg t_c$, to reach thermal equilibrium. Likewise, temperature control based on a feedback loop is inherently more time consuming. The proposed fast cooling should be advantageous for quantum information processing and storage based on quantum state distillation.

The findings that we report here establish a new link between frequent quantum measurements, which are the operational probes of short-time evolution, and non-equilibrium thermodynamical anomalies: heat and entropy rates of change that have the 'wrong' sign, rather than displaying their usual monotonic approach to equilibrium. These anomalies are determined by the oscillatory or negative values of the non-markovian quantum relaxation rates at short times corresponding to large energy uncertainty. They reveal unfamiliar general aspects of post-measurement quantum dynamics: the AZE, which was initially proposed as a means of enhancing or accelerating the initial-state change[12–15], here can either restore the equilibrium state or further depart from it via cooling. These anomalies underscore the fact that the system and the bath are inseparable (entangled)[6,23], even under weak-coupling conditions, a fact that has important implications for their short-time dynamics.

These results prompt further studies of a hitherto unexplored non-markovian short-time domain where neither the existing formulations of the second law of thermodynamics[2,5] nor the common notion that heat always flows from hotter to colder ensembles is applicable. This calls for an examination of non-markovian designs of quantum heat engines and their comparison to their markovian counterparts[24]. The short-time domain may also necessitate an in-depth scrutiny of fundamental quantum thermodynamical concepts. In particular, the need for temporal 'coarse graining' of entropy should be examined.

Received 18 October 2007; accepted 26 February 2008.

1. Landau, L. & Lifshitz, E. *Statistical Physics* 3rd edn Part 1 (Pergamon, 1980).
2. Spohn, H. Entropy production for quantum dynamical semigroups. *J. Math. Phys.* **19,** 1227–1230 (1978).
3. Alicki, R. The quantum open system as a model of the heat engine. *J. Phys. A* **12,** L103–L107 (1979).
4. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78,** 2690–2693 (1997).
5. Lindblad, G. Expectations and entropy inequalities for finite quantum systems. *Commun. Math. Phys.* **39,** 111–119 (1974).
6. Gelman, D. & Kosloff, R. Simulating dissipative phenomena with a random phase thermal wavefunctions, high temperature application of the Surrogate Hamiltonian approach. *Chem. Phys. Lett.* **381,** 129–138 (2003).
7. Zurek, W. H. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.* **75,** 715–775 (2003).
8. Braginsky, V. & Khalili, F. *Quantum Measurement* (Cambridge Univ. Press, Cambridge, 1995).
9. Barnett, S. Turning on quantum taps. *Nature* **362,** 113 (1993).
10. Haroche, S. & Raimond, J. *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, Oxford, 2006).
11. Misra, B. & Sudarshan, E. C. G. Zeno's paradox in quantum theory. *J. Math. Phys.* **18,** 756–763 (1977).
12. Lane, A. M. Decay at early times - larger or smaller than the golden rule. *Phys. Lett.* **99A,** 359–360 (1983).
13. Kofman, A. G. & Kurizki, G. Acceleration of quantum decay processes by frequent observations. *Nature* **405,** 546–550 (2000).
14. Facchi, P. & Pascazio, S. Quantum Zeno and inverse quantum Zeno effects. *Prog. Opt.* **42,** 147–217 (2001).
15. Kofman, A. G. & Kurizki, G. Unified theory of dynamically suppressed qubit decoherence in thermal baths. *Phys. Rev. Lett.* **93,** 130406 (2004).
16. Schulman, L. S. & Gaveau, B. Ratcheting up energy by means of measurement. *Phys. Rev. Lett.* **97,** 240405 (2006).
17. Allahverdyan, A. E. & Nieuwenhuizen, T. M. Extraction of work from a single thermal bath in the quantum regime. *Phys. Rev. Lett.* **85,** 1799–1802 (2000).
18. Allahverdyan, A. E., Gracià, R. S. & Nieuwenhuizen, T. M. Bath-assisted cooling of spins. *Phys. Rev. Lett.* **93,** 260404 (2004).
19. Cohen-Tannoudji, C., Dupont-Roc, J. & Grynberg, G. *Atom-Photon Interactions* (Wiley, New York, 1992).
20. Scully, M. O. Extracting work from a single thermal bath via quantum negentropy. *Phys. Rev. Lett.* **87,** 220601 (2001).
21. Beck, M., Jäckle, A., Worth, G. & Meyer, H.-D. The multiconfiguration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets. *Phys. Rep.* **324,** 1–105 (2000).
22. Nest, M. & Meyer, H. Dissipative quantum dynamics of anharmonic oscillators with the multiconfiguration time-dependent Hartree method. *J. Chem. Phys.* **119,** 24–33 (2003).
23. Stelmachovic, P. & Buzek, V. Dynamics of open quantum systems initially entangled with environment: Beyond the Kraus representation. *Phys. Rev. A* **64,** 062106 (2001).
24. Scully, M. O., Zubairy, M. S., Agarwal, G. S. & Walther, H. Extracting work from a single heat bath via vanishing quantum coherence. *Science* **299,** 862–864 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the support of DIP, GIF and EC (SCALA IP and MIDAS STREP).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.K. (gershon.kurizki@weizmann.ac.il).

LETTERS

Microscopic theory of the extraordinary optical transmission

Haitao Liu[1,2] & Philippe Lalanne[1]

The phenomenon of extraordinary light transmission through metallic films perforated by nanohole arrays at optical frequencies was first observed a decade ago[1] and initiated important further experimental and theoretical work. In view of potential applications of such structures—for example, subwavelength optics[2,3], optoelectronics devices[4,5], and chemical sensing[6]—it is important to understand the underlying physical processes in detail. Here we derive a microscopic theory of the transmission through subwavelength hole arrays, by considering the elementary processes associated with scattering of surface-plasmon-polariton (SPP) modes by individual one-dimensional chains of subwavelength holes. Using a SPP coupled-mode model that coherently gathers these elementary processes, we derive analytical expressions for all the transmission spectrum characteristics—such as the resonance wavelength, the peak transmission and the anti-resonance. Further comparisons of the model predictions with fully vectorial computational results allow us quantitatively to check the model accuracy and to discuss the respective impacts of SPP modes and of other electromagnetic fields on producing the extraordinary transmission of light. The model greatly expands our understanding of the phenomenon and may affect further engineering of nanoplasmonic devices.

Our present understanding of the extraordinary optical transmission (EOT) is based on the excitation of a surface electromagnetic Bloch mode at the front and rear metallic interfaces. The surface mode enhances the evanescent field at the hole apertures and thus contributes to funnelling the light through the hole array[6,7]. It exists in the long-wavelength regime[8,9] for almost-perfectly conducting metals, and has been recently studied at optical frequencies[10]. This well-accepted Bloch-mode picture represents a macroscopic description of the phenomenon, because it relies on a collective electromagnetic property of the metal interface perforated by a two-dimensional (2D) array of infinitely deep holes. Here we aim instead at providing a microscopic description. To do so, we abandon the classical mode-expansion approach[7–10], and consider the elementary scattering events occurring among the individual hole chains of the 2D array. Through a SPP coupled-mode model, we show how the coherent association of all these events builds up resonances in the 2D array. The model both corroborates the classical mode picture and provides a comprehensive framework that explains the physical origin of the Bloch mode. The EOT can thus be understood as resulting from a subtle surface-wave dynamical scattering process on the film interfaces. The dynamics involves two fields, the SPP mode of the flat interface and an additional field with a quasi-cylindrical behaviour.

At a microscopic level, the basic mechanism enabling the EOT is a coherent diffraction by all the individual holes acting as elementary scatterers[11]. Here we instead consider a single one-dimensional array of holes (a periodic hole chain with periodicity a_y in the y direction) perforating a metal substrate to be the elementary scatterers. This choice leads to a conceptually simple and explicit multiple-scattering SPP formalism that provides closed-form expressions for the EOT, and thus that allows us to analyse the phenomenon in depth.

The elementary scattering events we consider are shown in Fig. 1a–c for non-conical diffraction geometries (the y component k_y of the in-plane wavevector momentum is null). On interaction with the chain, the SPP modes are partly excited, transmitted, reflected or scattered into the chain and into a continuum of outgoing plane waves. The interaction defines six elementary scattering coefficients that we have calculated (see details in the Supplementary Information) by using a fully vectorial method[12] and the Lorentz reciprocity theorem[13]. We focus on two of the six coefficients—the SPP modal reflectance and transmittance coefficients ρ and τ. Their spectra are shown in Fig. 1e for near-infrared frequencies. For the computation, we have considered a real lossy metal, gold with a frequency-dependent permittivity of ε_g (where g refers to gold) given by the tabulated data in ref. 14. All six elementary scattering coefficients exhibit a smooth spectral behaviour, and only their associations through multiple scattering may result in anomalies or resonances.

Starting with only the knowledge of the elementary-event scattering coefficients in Fig. 1a–c, we derived a 'pure' SPP coupled-mode model that provides closed-form expressions for the optical transmittance and reflectance of arrayed geometries formed by a finite (or infinite) number of chains placed at arbitrary positions in the x direction. By 'pure' SPP model, we mean that the electromagnetic interaction among the chains is mediated only by SPP modes of flat interfaces, all other scattered waves being neglected (details of the SPP model are in the Supplementary Information). Here, we restrict ourselves to infinite and fully periodic structures (Fig. 1d) with the same periodicities in the x and y directions ($a_x = a_y = a$), in which the EOT was first observed in the near-infrared[1], and we consider only the case of normal incidence for the sake of simplicity. The main strength of the model is to provide closed-form expressions for the transmittance and reflectance coefficients $t_A(k_x)$ and $r_A(k_x)$ (where A refers to array) of the fundamental Bloch mode supported by the 2D hole array at the front and rear interfaces (Fig. 1d). For normal incidence ($k_x = 0$), these coefficients are simply given by

$$t_A(k_x = 0) = t + \frac{2\alpha\beta}{u^{-1} - (\rho + \tau)} \quad (1)$$

$$r_A(k_x = 0) = r + \frac{2\alpha^2}{u^{-1} - (\rho + \tau)} \quad (2)$$

where t and β denote $t(k_x = 0)$ and $\beta(k_x = 0)$, respectively. From $t_A(k_x)$ and $r_A(k_x)$, the zeroth-order transmittance coefficient $t_F(k_x)$ of the membrane is straightforwardly obtained by the Fabry–Pérot

[1]Laboratoire Charles Fabry de l'Institut d'Optique, CNRS, Univ. Paris-Sud, Campus Polytechnique, RD 128, 91127 Palaiseau cedex, France. [2]Key Laboratory of Opto-electronic Information Science and Technology, Ministry of Education, Institute of Modern Optics, Nankai University, Tianjin 300071, China.

equation (see Supplementary Information). In equations (1) and (2), $u = \exp(ik_{SPP}a)$ is the phase delay accumulated by the SPP over a grating period, $k_{SPP} = k_0[\varepsilon_g/(\varepsilon_g+1)]^{1/2}$ being the complex propagation constant of a SPP on a flat gold–air interface. As discussed below, the interaction between adjacent hole chains is not solely driven by SPP modes, which raises the question of to what extent a pure SPP model can capture the EOT phenomenon.

To answer this, we compare the model predictions with fully vectorial computational results obtained with a grating solver that relies on the rigorous coupled-wave analysis (RCWA). Figure 2 summarizes the main results of the inter-comparison obtained for a 2D hole array in a self-suspended gold membrane in air under transverse magnetic polarized illumination at near-infrared frequencies. On a linear scale, Fig. 2a shows the (0,0)-order transmittance T as a function of the frequency and of the in-plane wavevector k_x, and Fig. 2b and c shows typical zeroth-order transmittance $T(\lambda) = |t_F(k_x)|^2$ (where $k_x = (2\pi/\lambda)\sin\theta$ is dependent on wavelength λ for a fixed θ) and reflectance $R(\lambda)$ spectra for two incidence angles, $\theta = 0^\circ$ and 5°. There are differences between the model predictions and the RCWA data. Nevertheless, the SPP model is found to capture quantitatively the main features of the calculated data: the existence of two transmission branches in the (ω, k_x) plane (where ω is the light frequency) (Fig. 2a), the broadband lineshape of the EOT (Fig. 2b), the anti-resonance transmission dips located on the high-energy side of each resonance (inset in Fig. 2b), and the reflectance resonance the apparent dips of which are located at almost the same energies as those of the transmission peaks (Fig. 2c). This is all also consistent with the experimental results[1,15,16].

The microscopic SPP model shines new light on how the EOT resonance peaks are built up. From equations (1) and (2), the resonances originate from the presence of zeroes in the multiple-scattering denominator $u^{-1} - (\rho + \tau)$. The latter involves only the two elementary elastic scattering coefficients, ρ and τ. Two routes of loss are available, scattering into the continuum of radiation plane waves and scattering into the chain modes. Free-space radiation is kept low, especially for tiny holes, and because the hole modes are all below the cut-off frequencies of propagative modes and thus do not carry energy, the SPP scattering process of Fig. 1a is almost energy conservative. In fact $|\rho + \tau| \approx |\tau|$ is very close to one, and thus to $|u^{-1}|$, which is slightly larger than unity because of the small SPP damping through propagation over the subwavelength period. Thus the denominators of t_A and r_A can be made to be very close to zero whenever the SPP scattering events of the hole chains constructively interfere for

$$\mathrm{Re}(k_{SPP})a_x + \arg(\rho + \tau) \approx \mathrm{Re}(k_{SPP})a_x + \arg(\tau) = 0 \text{ modulo } 2\pi \quad (3)$$

where the function 'arg' refers to argument. The phase-matching condition of equation (3) and its effect on producing the transmission peaks in the EOT are detailed in the Supplementary Information. For the conditions for resonance, we emphasize that $\arg(\tau)$ is slightly positive but not null (Fig. 1e). This is because, when flying over the hole chain, the SPP mode experiences an artificial metal with a slightly lower effective conductivity ($n_{eff}^2 \approx f\varepsilon_{air} + (1-f)\varepsilon_g$, where $f = D/a_y$), and accumulates a weak phase delay, $\arg(\tau)$, approximately given by $[k_0\mathrm{Re}(n_{eff}) - \mathrm{Re}(k_{SPP})]D$, where D is the hole side length. This positive phase delay explains well why the front and rear interfaces of the perforated membrane resonate for wavelengths slightly larger than $\lambda_{SPP} = a_x[\varepsilon_g/(\varepsilon_g+1)]^{1/2}$, the wavelength for which a SPP mode on a flat interface is matched to the incident light through the reciprocal grating momentum: $\mathrm{Re}(k_{SPP})a_x = 0$ modulo 2π. We note that for $\lambda = \lambda_{SPP}$, the RCWA data and the model predictions both predict that $|t_A|$ is very weak, as shown in the inset in Fig. 2b.

Although we have focused here on tiny holes only, we note that the elementary scattering process in Fig. 1a is no longer energy conservative for larger holes (or slits) that sustain a propagative (non-evanescent) mode, because a large fraction of the incident SPP is then scattered into the hole mode[17]. Therefore τ becomes distinctly smaller than unity, and the resonance condition is no longer described by

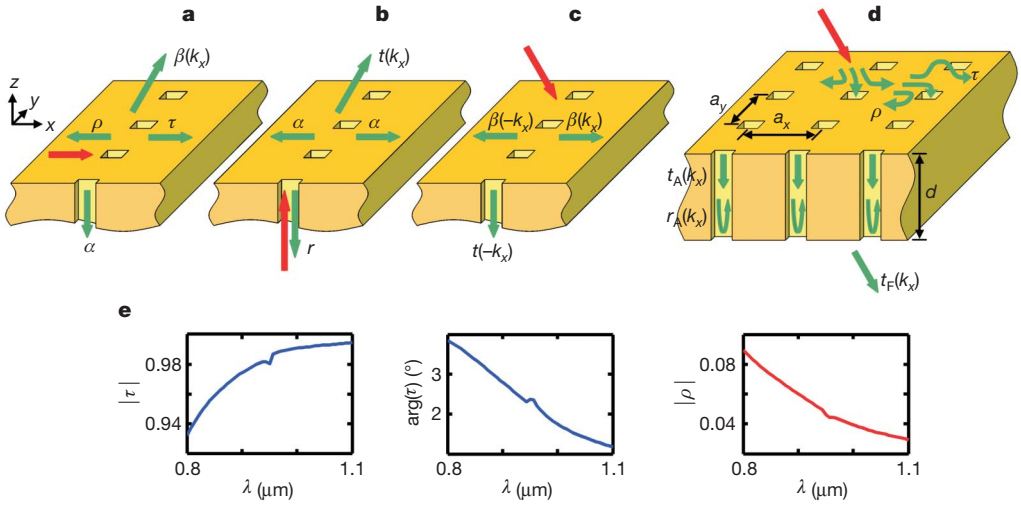


Figure 1 | Elementary processes involved in the EOT. They are all associated with the scattering of an electromagnetic field by a one-dimensional hole chain under illumination by the SPP mode (**a**), the fundamental Bloch mode of the hole chain (**b**), and an incident transverse magnetic polarized (magnetic vector along y axis) plane wave impinging with an oblique incidence defined by its in-plane wavevector component k_x (**c**). The red and green arrows refer to the incident and scattered modes, respectively. The processes in **a**, **b** and **c** define six independent elementary scattering coefficients: ρ (the reflectance coefficient of the SPP mode), τ (the transmittance coefficient of the SPP mode), α (the scattering coefficient from the SPP mode to the fundamental Bloch mode and vice versa according to the reciprocity theorem), $\beta(k_x)$ (the scattering coefficient from the SPP mode to the outgoing plane wave with an in-plane wavevector component k_x), $t(k_x)$ (the scattering coefficient from the fundamental Bloch mode to the plane wave) and r (the reflectance coefficient of the fundamental Bloch mode). Through the SPP model, these elementary scattering coefficients are combined to derive closed-form expressions for the transmittance and reflectance coefficients $t_A(k_x)$ and $r_A(k_x)$ of the fundamental Bloch mode supported by the 2D gold membrane shown in **d**, which then give the (0,0)-order transmittance coefficient $t_F(k_x)$ of the membrane by the classical Fabry–Pérot equation. (d is the membrane thickness.) Of the six coefficients, the elastic scattering coefficients, ρ and τ, are fundamental in setting up the EOT. Their calculated moduli and the argument of τ are shown in **e** at optical frequencies for $a_y = 940$ nm and for square holes with a side length of $D = 266$ nm.

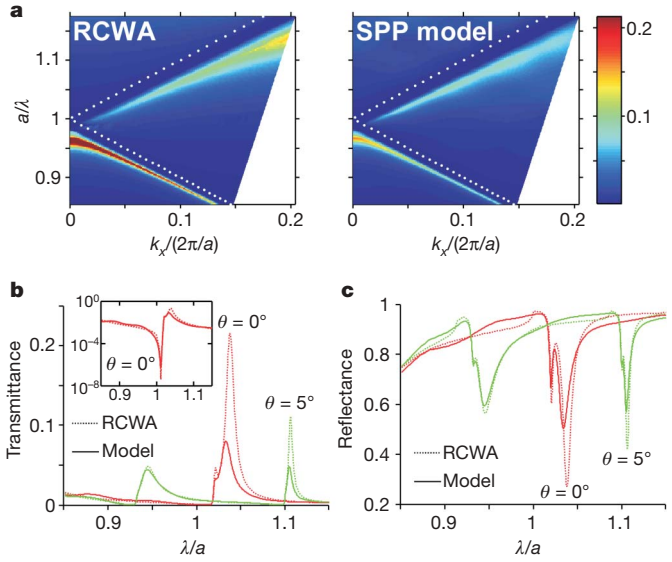


Figure 2 | Comparison between the SPP model predictions and fully vectorial (RCWA) computation data obtained for the EOT of a gold hole-array membrane in air. The SPP model quantitatively captures all the salient features of the EOT, such as the existence of two transmission branches in the (ω, k_x) plane (**a**), the broadband lineshape of the EOT (**b**), the anti-resonance transmission dips (inset to **b**), and the reflectance spectra (**c**). **a**, Colour-scale images showing the (0,0)th-order far-field transmittance T as a function of the frequency and of the in-plane wavevector k_x. The white dotted curves correspond to the air light lines (that is, the positions in the (ω, k_x) plane where diffracted orders become tangent to the surface). The scale is linear and the maximum EOT is ~21%. **b**, Transmittance spectra for two angles of incidence, $\theta = 0^\circ$ and 5°. The inset, showing the transmission dips, is plotted using a log scale. **c**, Corresponding reflectance spectra. All the results are obtained for a gold membrane (in air) perforated by a 2D hole array with $a_x = a_y = a = 940$ nm, $D = 266$ nm and membrane thickness $d = 200$ nm (hole filling fraction of 8%). In **b** and **c**, the dots and solid curves correspond to RCWA data and to SPP model predictions, respectively.

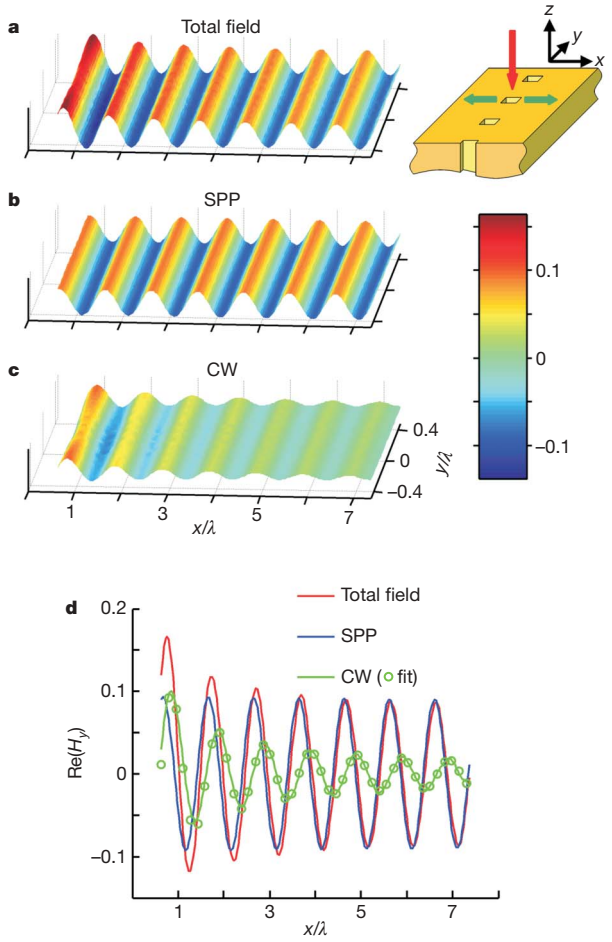


Figure 3 | Surface waves generated on a gold surface by a single hole chain, illuminated by a normally incident plane wave polarized along the *x* axis with a unitary magnetic field at the gold surface. **a**, 2D map of the real part of the total H_y-component field scattered in the x–y plane. A single period in the y direction is shown. For clarity, the incident and specular-reflected plane waves have been removed. The inset to **a** shows a single hole chain. **b**, **c**, Extracted SPP-mode and cylindrical-wave contributions. **d**, Cross-sectional representation ($y = a_y/2$) showing the damping characteristic lengths of the total-field (red), the SPP-mode (blue) and the cylindrical-wave (green) contributions. The green circles represent a function of type $x^{-0.84}\cos(k_0 x + \varphi)$, which is fitted from the quasi-cylindrical-wave (CW) data. The computational results are obtained for $a_y = 940$ nm and $D = 266$ nm, and for an illumination at $\lambda = 974$ nm.

equation (3), but by the Fabry–Pérot resonance condition of the hole-array Bloch modes[10].

Although it captures the main features of the EOT at near-infrared frequencies well, the multiple-scattering SPP model does not provide a complete microscopic picture. For instance, it accounts for only half of the total transmitted energy at peak transmittance (Fig. 2b). To analyse this, we calculated the field scattered by a single hole chain on the metal interface under illumination by a normally incident plane wave. The real part of the y component H_y of the total magnetic-field vector is shown in Fig. 3a as a function of the in-plane x and y coordinates. We extracted the SPP-mode contribution $\beta\exp(ik_{SPP}x)$ (Fig. 3b), and by taking the difference from the total field, the remaining contribution (Fig. 3c). As expected[18–23], the total nearby field scattered on the interface by the nanohole array is not a pure SPP mode; it additionally incorporates a quasi-cylindrical wave creeping along the interface over several wavelength distances, with an amplitude damping that scales approximately as $(1/x)^{1/2}$. At $\lambda = 0.974$ μm, it is found that the cylindrical-wave-field decay rate is much faster than that of the SPP, but at subwavelength distances ($x/\lambda \approx 1$) from the chain (like those encountered in 2D nanohole arrays), the two waves almost equally contribute to the total field. The two waves belong to the same scattering problems, so they are additionally excited with a moderate phase difference, and because they propagate with similar phase velocities, $k_{SPP} \approx k_0$, their contributions to the EOT sum up almost constructively.

The previous results hold in the near-infrared band, but the EOT has also been observed in other spectral ranges[1,24,25]. We performed similar analyses at visible and thermal infrared frequencies. It turns out that the SPP model becomes less and less accurate as the metal conductivity increases. The physics is very similar to that recently discussed for slits[20]. At thermal infrared wavelengths, the SPP is only weakly excited by the hole chains and the multi-scattering process is dominantly driven by the cylindrical wave, which is equally excited at all energies (Supplementary Information). Although more studies are needed, we believe that the present SPP model clarifies the physics of the EOT phenomenon well on the microscopic scale and that it may be helpful for further engineering of nanoplasmonic devices relying for instance on high SPP-excitation probabilities. The dual-wave picture arising from the SPP model may additionally reconcile earlier views[1,18] on the role of SPP in the EOT. However, we note that the quasi-cylindrical wave is conceptually different[20,21,23] from the wave model in refs 18 and 19.

Received 27 September 2007; accepted 24 January 2008.

1. Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* **391**, 667–669 (1998).

2. Alkaisi, M. M., Blaikie, R. J., McNab, S. J., Cheung, R. & Cumming, D. R. S. Sub-diffraction-limited patterning using evanescent near-field optical lithography. *Appl. Phys. Lett.* **75,** 3560–3562 (1999).
3. Luo, X. G. & Ishihara, T. Sub-100-nm photolithography based on plasmon resonance. *Jap. J. Appl. Phys.* **43,** 4017–4021 (2004).
4. Collin, S., Pardo, F. & Pelouard, J. L. Resonant-cavity-enhanced subwavelength metal-semiconductor-metal photodetector. *Appl. Phys. Lett.* **83,** 1521–1523 (2003).
5. Liu, C., Kamaev, V. & Vardeny, Z. V. Efficiency enhancement of an organic light emitting diode with a cathode forming two-dimensional periodic hole array. *Appl. Phys. Lett.* **86,** 143501 (2005).
6. Genet, C. & Ebbesen, T. W. Light in tiny holes. *Nature* **445,** 39–46 (2007).
7. Martin-Moreno, L. *et al.* Theory of extraordinary optical transmission through subwavelength hole arrays. *Phys. Rev. Lett.* **86,** 1114–1117 (2001).
8. Pendry, J. B., Martin-Moreno, L. & Garcia-Vidal, J. F. Mimicking surface plasmons with structured surfaces. *Science* **305,** 847–848 (2004).
9. García de Abajo, F. J. & Sáenz, J. J. Electromagnetic surface modes in structured perfect-conductor surfaces. *Phys. Rev. Lett.* **95,** 233901 (2005).
10. Lalanne, P., Rodier, J. C. & Hugonin, J. P. Surface plasmons of metallic surfaces perforated by nanohole arrays. *J. Opt. Pure Appl. Opt.* **7,** 422–426 (2005).
11. Genet, C., van Exter, M. P. & Woerdman, J. P. Huygens description of resonance phenomena in subwavelength hole arrays. *J. Opt. Soc. Am. A* **22,** 998–1002 (2005).
12. Silberstein, E., Lalanne, P., Hugonin, J. P. & Cao, Q. On the use of grating theory in integrated optics. *J. Opt. Soc. Am. A* **18,** 2865–2875 (2001).
13. Snyder, A. W. & Love, J. D. *Optical Waveguide Theory* 602–608 (Chapman and Hall, London/New York, 1983).
14. Palik, E. D. *Handbook of Optical Constants of Solids* Part II (Academic, New York, 1985).
15. Barnes, W. L., Murray, W. A., Dintinger, J., Devaux, E. & Ebbesen, T. W. Surface plasmon polaritons and their role in the enhanced transmission of light through periodic arrays of subwavelength holes in a metal film. *Phys. Rev. Lett.* **92,** 107401 (2004).
16. Ye, Y.-H. & Zhang, J.-Y. Middle-infrared transmission enhancement through periodically perforated metal films. *Appl. Phys. Lett.* **84,** 2977–2979 (2004).
17. Lalanne, P., Hugonin, J. P. & Rodier, J. C. Theory of surface plasmon generation at nanoslit apertures. *Phys. Rev. Lett.* **95,** 263902 (2005).
18. Lezec, H. J. & Thio, T. Diffracted evanescent wave model for enhanced and suppressed optical transmission through subwavelength hole arrays. *Opt. Express* **12,** 3629–3641 (2004).
19. Gay, G. *et al.* The optical response of nanostructured surfaces and the composite diffracted evanescent wave model. *Nature Phys.* **2,** 262–267 (2006).
20. Lalanne, P. & Hugonin, J. P. Interaction between optical nano-objects at metallo-dielectric interfaces. *Nature Phys.* **2,** 551–556 (2006).
21. Aigouy, L. *et al.* Near-field analysis of surface waves launched at nano-slit apertures. *Phys. Rev. Lett.* **98,** 153902 (2007).
22. Boersma, J. & Lee, S. W. An exact solution for diffraction of a line-source field by a half-plane. *J. Math. Phys.* **18,** 321–328 (1977).
23. García-Vidal, F. J., Rodrigo, S. G. & Martín-Moreno, L. Foundations of the composite diffracted evanescent wave model. *Nature Phys.* **2,** 790 (2006).
24. Gomez Rivas, J., Schotsch, C., Haring Bolivar, P. & Kurz, H. Enhanced transmission of THz radiation through subwavelength holes. *Phys. Rev. B* **68,** 201306(R) (2003).
25. Shou, X., Agrawal, A. & Nahata, A. Role of metal thickness on the enhanced transmission properties of a periodic array of subwavelength apertures. *Opt. Express* **13,** 9834–9840 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements H.L. acknowledges a fellowship of the ''Fondation Franco-Chinoise pour la Science et ses Applications'' (FFCSA) and the China Scholarship Council (CSC). We thank J. P. Hugonin for discussions and computational assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.L. (philippe.lalanne@institutoptique.fr).

Vol 452|10 April 2008|doi:10.1038/nature06817

LETTERS

Improper ferroelectricity in perovskite oxide artificial superlattices

Eric Bousquet[1]*, Matthew Dawber[2]*†, Nicolas Stucki[2], Céline Lichtensteiger[2], Patrick Hermet[1], Stefano Gariglio[2], Jean-Marc Triscone[2] & Philippe Ghosez[1]

Ferroelectric thin films and superlattices are currently the subject of intensive research[1,2] because of the interest they raise for technological applications and also because their properties are of fundamental scientific importance[3–5]. Ferroelectric superlattices[6] allow the tuning of the ferroelectric properties while maintaining perfect crystal structure and a coherent strain, even throughout relatively thick samples. This tuning is achieved in practice by adjusting both the strain[7–10], to enhance the polarization, and the composition, to interpolate between the properties of the combined compounds[11–15]. Here we show that superlattices with very short periods possess a new form of interface coupling, based on rotational distortions, which gives rise to 'improper' ferroelectricity. These observations suggest an approach, based on interface engineering, to produce artificial materials with unique properties. By considering ferroelectric/paraelectric $PbTiO_3$/$SrTiO_3$ multilayers, we first show from first principles that the ground-state of the system is not purely ferroelectric but also primarily involves antiferrodistortive rotations of the oxygen atoms in a way compatible with improper ferroelectricity. We then demonstrate experimentally that, in contrast to pure $PbTiO_3$ and $SrTiO_3$ compounds, the multilayer system indeed behaves like a prototypical improper ferroelectric and exhibits a very large dielectric constant of $\varepsilon_r \approx 600$, which is also fairly temperature-independent. This behaviour, of practical interest for technological applications[16], is distinct from that of normal ferroelectrics, for which the dielectric constant is typically large but strongly evolves around the phase transition temperature and also differs from that of previously known improper ferroelectrics that exhibit a temperature-independent but small dielectric constant only.

In an artificially layered ferroelectric–dielectric superlattice, electrostatic coupling between alternating thin ferroelectric and dielectric layers is able to induce a polarization in the latter[11]. When the dielectric is sufficiently polarizable, this yields a uniformly and highly polarized ground state[11,14], the polarization of which is in general predictable by considering only the ferroelectric degree of freedom and simple electrostatic arguments[12,15]. In $PbTiO_3$/$SrTiO_3$ superlattices formed by the repetition of n_p unit cells of $PbTiO_3$ and n_s unit cells of $SrTiO_3$ (denoted n_p/n_s), the control of the $PbTiO_3$ volume fraction $n_p/(n_p + n_s)$ allows tuning of the polarization and phase transition temperature in a predictable way over a wide range of compositions[15]. However, the scaling law deduced from the usual arguments breaks down in the limit of ultrathin $PbTiO_3$ layers. We show here that this is due to the emergence of a new phenomenon: an interfacially induced form of improper ferroelectricity that gives rise to unique properties and presents an attractive pathway to new 'interfacially engineered' materials.

The reference cubic structure of ABO_3 perovskite compounds can be unstable to different kinds of energy-lowering distortions. In $PbTiO_3$, the cubic phase is unstable not only to a polar zone-centre distortion responsible for the ferroelectric (FE) ground state but also to a zone-boundary distortion involving tilts of the oxygen octahedra[17]. At the bulk level, the latter is suppressed when the polar distortion is condensed, but both can coexist at surfaces[18,19]. In bulk $SrTiO_3$, oxygen rotation is conversely responsible for a non-polar antiferrodistortive (AFD) ground state and ferroelectricity is suppressed by quantum fluctuations[20], but both distortions can coexist under pressure[20] or appropriate epitaxial strains[21–23], yielding complex phase diagrams. It is thus expected that ferroelectric and antiferrodistortive distortions will strongly compete in $PbTiO_3$/$SrTiO_3$ superlattices.

To determine theoretically the ground-state structure and properties of such superlattices, we performed density functional theory calculations within the local density approximation. We used norm-conserving pseudopotentials and a plane-wave basis set, as implemented in the ABINIT package[24]. As in ref. 10, we adopted a supercell approach that (1) allows for the implicit treatment of the mechanical constraint imposed by the substrate by fixing the in-plane lattice constant of the superlattice, and (2) imposes short-circuit electrical boundary conditions through the use of periodic boundary conditions. However, we additionally doubled the size of the supercell in-plane to allow for antiferrodistortive oxygen motions[22]. Starting from the prototype paraelectric space-group *P*4/*mmm* structure corresponding to the highest achievable symmetry, we identified instabilities from the inspection of the phonon dispersion curves and accordingly lowered the symmetry and performed new structural relaxations. This was repeated until no instabilities were present.

For the simulation of a $PbTiO_3$/$SrTiO_3$ 1/1 superlattice grown on a [001] $SrTiO_3$ substrate, inspection of the phonon dispersion curves of the *P*4/*mmm* phase reveals the existence of zone-centre ferroelectric unstable modes with polarization *P* out-of-plane (Γ_3^- mode[25], Fig. 1a) or in-plane (Γ_5^- mode), called respectively FE_z and FE_{xy}, where the z and xy indices refer to out-of-plane and in-plane spatial directions respectively. Moreover, significantly larger antiferrodistortive instabilities are also present at the M (1/2, 1/2, 0) point, which correspond to different kinds of tilts of the oxygen octahedra. This includes both tilts around the [001] axis with successive octahedra along [001] moving out-of-phase (M_4^- mode, Fig. 1b, called AFD_{zo}) or in-phase (M_2^+ mode, Fig. 1c, called AFD_{zi}), and tilts around an axis perpendicular to (001) (M_5^- mode, called AFD_{xy}). Considering the sublattice of oxygen octahedra only, AFD_{zi} and AFD_{zo} are respectively equivalent to $a^0a^0c^+$ and $a^0a^0c^-$ in Glazer's notation[26].

We performed structural optimizations under symmetry constraints compatible with the condensation of individual or coupled

[1]Physique Théorique des Matériaux, Université de Liège, Allée du 6 Août 17 (B5), 4000 Sart Tilman, Belgium. [2]DPMC, University of Geneva, 24 Quai E.-Ansermet 1211, Geneva 4, Switzerland. †Present address: Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-3800, USA.
*These authors contributed equally to this work.

unstable modes identified above for various epitaxial strains, yielding the energy diagram of Fig. 2a. Inspection of the curves shows that the ground state is strongly dependent on the epitaxial strain and always involves the coupling of ferroelectric and antiferrodistortive distortions.

At the epitaxial strain corresponding to a $SrTiO_3$ substrate ($SrTiO_3$ misfit strain equal to zero), Fig. 2a shows that allowing a FE_z distortion alone would produce a negligible decrease of energy, whereas much more energy is gained when we allow for the AFD_{zo} distortion. Condensation of the AFD_{zo} distortion suppresses the individual FE_z and AFD_{zi} instabilities in the resulting *P*4/*nbm* phase but, surprisingly, the energy can be lowered further by allowing both the AFD_{zi} and FE_z distortions to coexist with the AFD_{zo} distortion. Proper energy minimization thus yields an initially unexpected $(AFD/FE)_z$ ground state combining FE_z, AFD_{zi} and AFD_{zo} distortions. This ground state exhibits a spontaneous polarization of 26 $\mu C\,cm^{-2}$ and oxygen rotation angles of $+6.2^\circ$ and -1.8° in alternating layers ($\phi_{zi} \pm \phi_{zo}$ from Fig. 1 and Fig. 2b).

Calculations at other thicknesses (3/3, 5/3, 7/3) reveal the existence of a similar antiferrodistortive/ferroelectric ground state. In all cases, the polarization is nearly homogeneous through the whole supercell, which is electrostatically required to avoid large depolarizing fields[11]. However, although oxygen tilts of similar amplitude are present at all thicknesses in the interface unit cell, these tilts rapidly disappear in the interior of the layers, revealing that the antiferrodistortive/ferroelectric coupling is essentially an interfacial effect. Its origin can be traced back to the asymmetric environment of the interfacial oxygen atoms. From the inspection of the interatomic force constants, it appears that the atoms in the rotating oxygen octahedra at the interface interact differently with up and down Sr and Pb atoms. This produces a net coupling between the cation FE_z motion and the oxygen in-plane AFD_{zi} motion that explains the observed $(FE/AFD)_z$ ground state in the superlattice. The coupling is absent in parent bulk compounds where the up and down cations are identical, their interaction with the oxygen atoms thus cancelling out.

As is also shown in Fig. 2c, over a wide range of epitaxial strains, the antiferrodistortive/ferroelectric coupling significantly enhances the spontaneous polarization compared to that of a purely FE_z state. Moreover, it can produce a sizeable polarization in cases where it would not be observed otherwise. All this strongly suggests that the unexpected recovery of ferroelectricity in $PbTiO_3/SrTiO_3$ multilayers experimentally reported[15] in the limit of ultrathin $PbTiO_3$ layers originates from the coupling of ferroelectric and antiferrodistortive distortions.

The ground state of short-period superlattices is not purely ferroelectric, so we can expect a phase transition mechanism distinct from that of ordinary ferroelectrics. In the 1/1 superlattice on $SrTiO_3$ for instance, the ground state results from the condensation of three distortions—FE_z, AFD_{zi} and AFD_{zo}—related to independent order parameters P_z, ϕ_{zi} and ϕ_{zo}. Because the FE_z instability is virtually suppressed, P_z will certainly not appear as the primary order parameter. As previously discussed by Levanyuk[27] and Holakovski[28], different phase-transition mechanisms (such as improper or triggered ferroelectricity) can be observed when independent order parameters are involved; the exact nature of the transition depends on the way order parameters are allowed to couple into a Landau-type free energy expansion. Typical improper ferroelectrics with primary order parameter ϕ_{zi} and ϕ_{zo} exhibit a generic Landau expansion which includes a term scaling linearly with the polarization[27]:

$$\Phi = (T - T_C)\left(a_{zi}\phi_{zi}^2 + a_{zo}\phi_{zo}^2\right) + \alpha P_z^2 - g\phi_{zi}\phi_{zo}P_z + 4\text{th order terms} \quad (1)$$

Here T_C is the transition temperature related to the primary order parameter and a_{zi}, a_{zo}, α and g are temperature-independent parameters. From the inspection of the irreducible representations of FE_z (Γ_3^-), AFD_{zi} (M_2^+) and AFD_{zo} (M_4^-) in the *P*4/*mmm* phase we can see that the product $\phi_{zi}\phi_{zo}$ transforms as P_z (see Supplementary Information) so that, in contrast to ordinary ferroelectrics that allow only even-power terms in P_z, the lowest-order P_z invariant term in the Landau expansion is linear and has the form $-g\phi_{zi}\phi_{zo}P_z$. Calculation of the energy for the 1/1 superlattice as a function of P_z around $P_z = 0$ with fixed ϕ_{zi} and ϕ_{zo} (shown in Fig. 2d) not only demonstrates the existence of such a linear term in P_z but also highlights the sizeable value of the coupling parameter g (see also Supplementary Information).

From equation (1), the existence of this coupling term between the distortions is a mandatory but not totally sufficient requirement for improper ferroelectricity, which additionally requires that ϕ_{zi} and ϕ_{zo} have identical transition temperature. Although this is automatically fulfilled in prototypical examples of improper ferroelectrics in which ϕ_{zi} and ϕ_{zo} are components of a two-dimensional order parameter, it

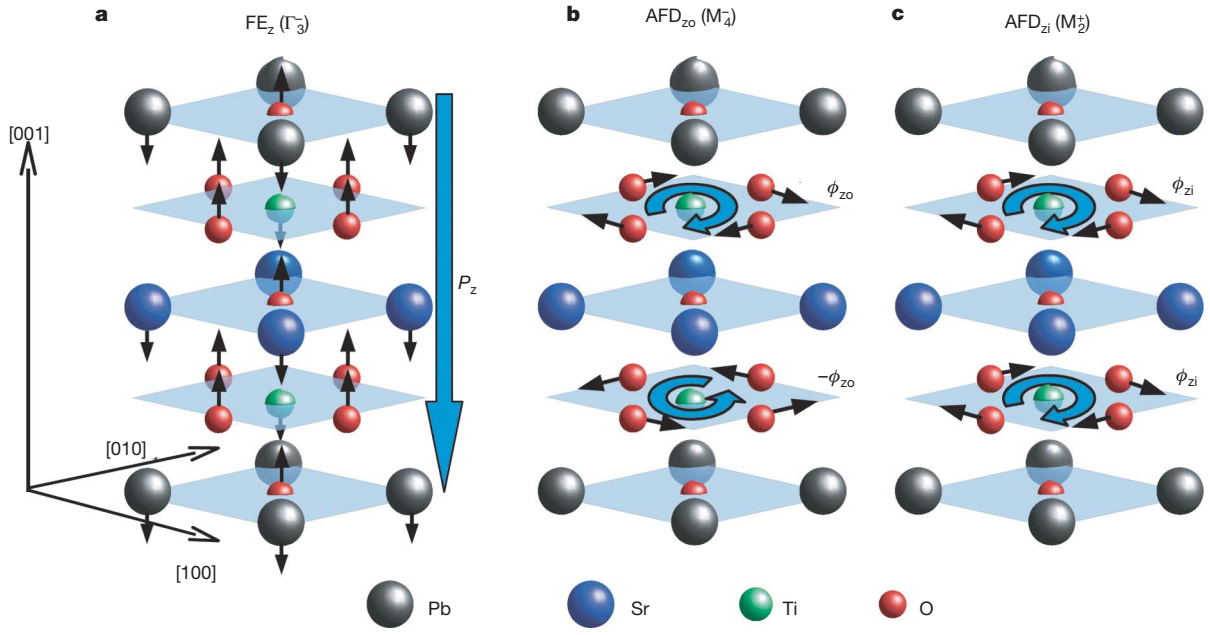


Figure 1 | Schematic view of the prototype *P*4/*mmm* unit cell of the 1/1 $PbTiO_3/SrTiO_3$ superlattice and atomic motions associated to different energy lowering distortions. **a**, FE_z (Γ_3^- mode) giving rise to a polarization P_z. **b**, AFD_{zo} (M_4^- mode) with oxygen rotation angle ϕ_{zo}. **c**, AFD_{zi} (M_2^+ mode) with oxygen rotation angle ϕ_{zi}.

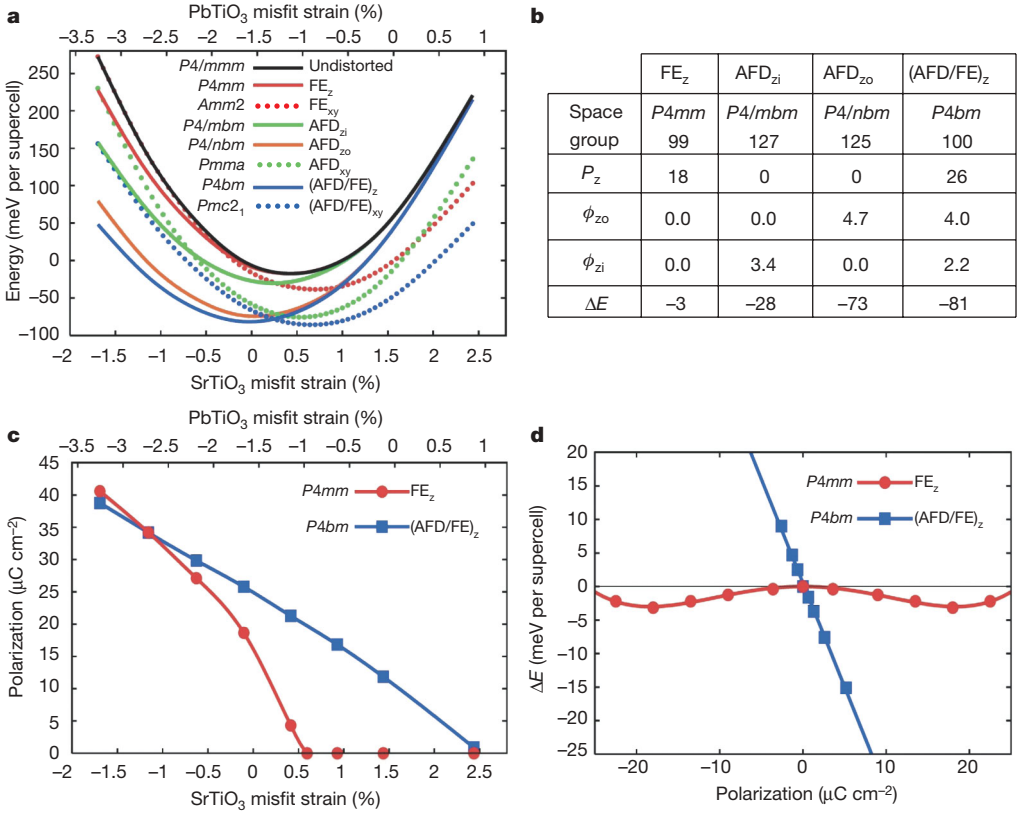
cannot be unambiguously inferred here from theoretical arguments because ϕ_{zi} and ϕ_{zo} belong to different one-dimensional irreducible representations. It is, however, achieved in practice, which is demonstrated experimentally below from the observation of a unique T_C and a typical improper ferroelectric behaviour of the material properties with temperature.

Improper ferroelectrics show critical behaviour different to that of ordinary ferroelectrics and the fingerprints of improper ferroelectricity can be tracked in the distinct temperature dependencies of the polarization and dielectric constant. Therefore we experimentally compared the behaviour of $PbTiO_3/SrTiO_3$ superlattices corresponding to the two regimes identified in ref. 12. The first regime corresponds to a range of compositions and thicknesses producing ordinary behaviour (such as a 9/3 sample) and the second regime corresponds to superlattices exhibiting an anomalous recovery of ferroelectricity (such as a 2/3 or 1/3 sample).

These samples were prepared by off-axis magnetron sputtering on (001) $SrTiO_3$ single-crystal substrates with TiO_2 termination. In addition to the superlattices, high-quality epitaxial top and bottom $SrRuO_3$ electrodes (important for the accurate determination of electrical properties) were deposited *in situ*. X-ray diffraction confirmed that the entire structure was epitaxially constrained in-plane to the $SrTiO_3$ lattice constant, and atomic force microscopy analyses showed that the unit-cell steps originally present on the substrate were carried through to the surface of the structures (see ref. 12 for more details on the sputtering process and growth conditions).

In $PbTiO_3/SrTiO_3$ superlattices, we have previously shown that epitaxial strain makes the ferroelectric phase transition second order[15]. If the system behaves like an ordinary mean-field ferroelectric, P_z is therefore expected from Landau theory to evolve like $(T_C - T)^{1/2}$ and such behaviour is indeed observed for the 9/3 sample, as shown in Fig. 3a. In improper ferroelectrics, however, the order parameter is no longer the polarization and it is instead ϕ_{zi} and ϕ_{zo} that, if they are mean-field, are expected to behave like $(T_C - T)^{1/2}$. From the linear P_z term in the Landau expansion, and if ϕ_{zi} and ϕ_{zo} exhibit similar T_C, this yields[27] $P_z \propto \phi_{zi}\phi_{zo} \propto (T_C - T)$, which is exactly what is observed for the 2/3 and 1/1 samples in Fig. 3b. Moreover, the experimentally measured polarization for the 1/1 sample extrapolates to 23 $\mu C\,cm^{-2}$ at zero temperature, in close agreement with the theoretical prediction of 26 $\mu C\,cm^{-2}$. The distinct evolution of the polarization in the two kinds of samples is further confirmed in the X-ray measurement of the tetragonality (that is, the ratio of the unit cell parameters c/a), which shows a linear dependence on $(T_C - T)$ for the 9/3 sample (Fig. 3c) and a quadratic dependence on $(T_C - T)$ for the 2/3 sample (Fig. 3d), in agreement with the behaviour expected from the strain-polarization coupling[17] (c/a scales as P_z^2).

In improper ferroelectrics, it is furthermore expected that the dielectric susceptibility will not obey a conventional Curie–Weiss law at the phase transition because the order parameter is no longer polar and will not directly couple with an external electric field. Instead, it can be shown[27] that the dielectric susceptibility will remain largely independent of temperature, with a step discontinuity at the transition temperature. Distinct behaviours are again coherently reported in Fig. 3e and f for the two kinds of samples. Whereas a typical Curie–Weiss law is observed for the 9/3 sample, the dielectric >constant of the 2/3 sample remains fairly independent over a wide range of temperature, except for a small step at the transition temperature, as expected for an improper ferroelectric transition.



	FE_z	AFD_{zi}	AFD_{zo}	$(AFD/FE)_z$
Space group	*P4mm* 99	*P4/mbm* 127	*P4/nbm* 125	*P4bm* 100
P_z	18	0	0	26
ϕ_{zo}	0.0	0.0	4.7	4.0
ϕ_{zi}	0.0	3.4	0.0	2.2
ΔE	–3	–28	–73	–81

Figure 2 | Results of the first-principles calculations. a, Relative stability, in terms of the misfit strain ($s = a/a_0 - 1$, where a_0 is the equilibrium cubic lattice constant), of different phases compatible with the condensation of individual or coupled instabilities in a $PbTiO_3/SrTiO_3$ 1/1 superlattice. **b**, Amplitude of the polarization P_z ($\mu C\,cm^{-2}$), rotation angles ϕ_{zi} and ϕ_{zo} (degrees) and energy gain ΔE with respect to the prototype *P4mmm* phase (meV per supercell) in different low-symmetry phases of a $PbTiO_3/SrTiO_3$ 1/1 superlattice epitaxially grown on $SrTiO_3$. **c**, Evolution of the spontaneous polarization with the misfit strain in the FE_z (*P4mm*) and $(AFD/FE)_z$ (*P4bm*) phases for a $PbTiO_3/SrTiO_3$ 1/1 superlattice epitaxially grown on $SrTiO_3$. **d**, Evolution of the energy with P_z in the FE_z phase ($\phi_{zi} = \phi_{zo} = 0$) and in the $(AFD/FE)_z$ phase ($\phi_{zi} = 2.2^\circ$ and $\phi_{zo} = 4.0^\circ$) for a $PbTiO_3/SrTiO_3$ 1/1 superlattice epitaxially grown on $SrTiO_3$.

Finally, we performed additional X-ray diffraction measurements on samples exhibiting typical improper ferroelectric behaviour (see Supplementary Information). These provided direct evidence of the unit cell doubling associated with oxygen rotations, definitively demonstrating the coherency between our experimental and theoretical findings.

The high values of the spontaneous polarization and dielectric constant of the 2/3 sample ($P_s = 11\ \mu C\,cm^{-2}$ and $\varepsilon_r \approx 600$ at 300 K) strongly contrast with what was previously reported for typical improper ferroelectrics[27] (in gadolinium molybdate, $P_s = 0.2\ \mu C\,cm^{-2}$ and $\varepsilon_r \approx 10$). In our short-period superlattices, although ferroelectricity is unambiguously improper, the polarization and dielectric constant are comparable to those in usual proper ferroelectrics owing to the intrinsically high polarizability of $PbTiO_3$ and $SrTiO_3$. The fact that as well as being large the dielectric constant is very stable over a wide range of temperature makes these artificial superlattices particularly attractive for dielectric applications.

The emergence of improper ferroelectricity as a direct product of the interfaces in the artificially layered structure also suggests a promising new approach in which 'interface engineering' is used to tune material properties. In our example, ferroelectric and antiferrodistortive instabilities were inherent to both $PbTiO_3$ and $SrTiO_3$ at the bulk level but the instabilities competed differently in each system and generated different ground states. Here we have shown how using artificial layering and interfaces can change the nature of the coupling between these instabilities and produce a new ground state and properties. The large variety of perovskite oxides exhibiting different tendencies to ferroelectric and antiferrodistortive instabilities suggest that there is tremendous potential to generate artificial materials with new behaviours in this way. Although we cannot provide a full set of 'design rules' at this stage, the method described here of combining compounds in which cations interact differently with oxygen atoms (such as Sr and Pb) is one way of creating interfacial asymmetry and thus modifying the ferroelectric/antiferrodistortive coupling.

We note that the 'interface engineering' approach complements rather than replaces other strategies to material tuning, such as strain engineering. Indeed, the role of epitaxial strain is important in our example. As shown in Fig. 2a, the strain can be used to switch from the $(AFD/FE)_z$ to the $(AFD/FE)_{xy}$ state. Especially intriguing is the equivalence in energy for these states at an applied strain of roughly 0.25%, which might allow considerable freedom for polarization rotation and a high piezoelectric response.

A clear benefit of the present approach, which uses ferroelectric/antiferrodistortive coupling instead of relying solely on the ferroelectric instability, is that it allows manipulation of the temperature dependence of the polar order parameter, which here leads to a high, nearly temperature-independent dielectric constant, a property

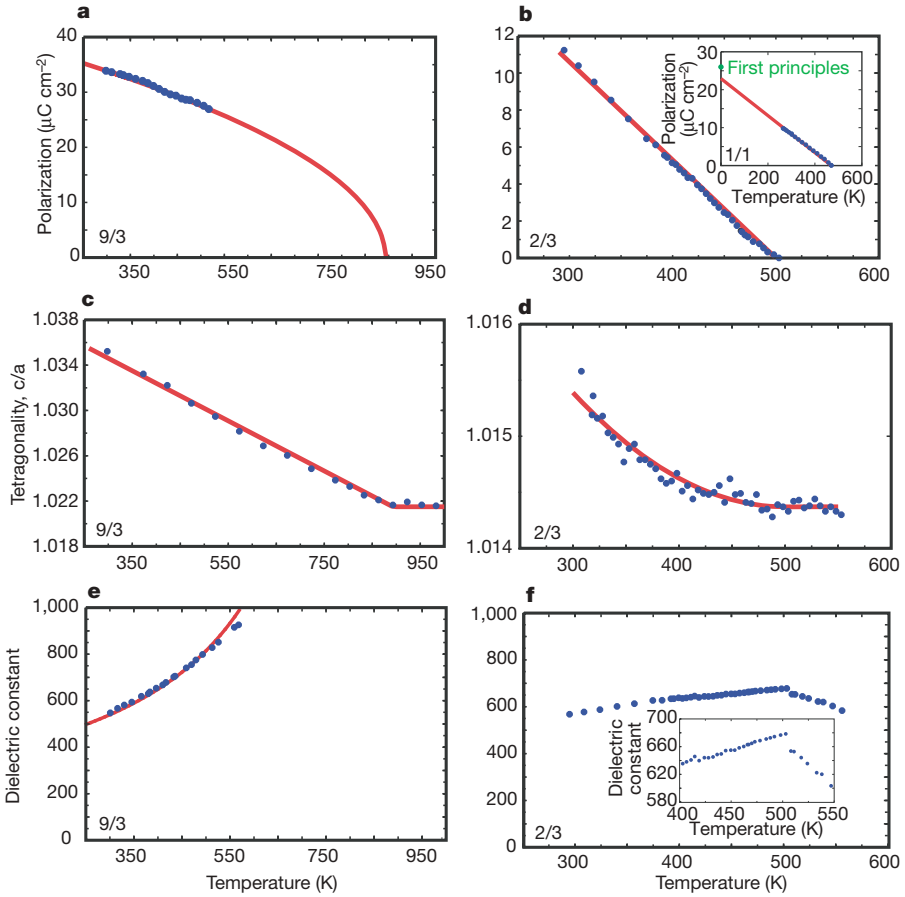


Figure 3 | Experimental measurements of phase transition behaviour in normal and anomalous samples. **a**, Ferroelectric polarization of a 100-nm-thick $PbTiO_3/SrTiO_3$ 9/3 superlattice as a function of temperature. The red line is a fit, $P_z = 1.39 \times (893 - T)^{1/2}$ (extrapolating to $P_z = 0$ at the transition temperature measured from **c**). **b**, Ferroelectric polarization of a 100-nm-thick $PbTiO_3/SrTiO_3$ 2/3 superlattice as a function of temperature. The red line is a fit, $P_z = 0.053 \times (500 - T)$. The inset shows the same measurement for a 100 nm $PbTiO_3/SrTiO_3$ 1/1 superlattice. **c**, Tetragonality of a 100-nm-thick $PbTiO_3/SrTiO_3$ 9/3 superlattice (measured using X-ray diffraction) as a function of temperature. The red line is a fit, $c/a = 1.0215 + 2.21 \times 10^{-5} \times (893 - T)$. **d**, Tetragonality of a 100-nm-thick $PbTiO_3/SrTiO_3$ 2/3 superlattice as a function of temperature. The red line is a fit, $c/a = 1.01437 + 2.39 \times 10^{-8} \times (500 - T)^2$. **e**, Dielectric constant of a 100-nm-thick $PbTiO_3/SrTiO_3$ 9/3 superlattice as a function of temperature. The red line is a fit, $\varepsilon = 320{,}000/(893 - T)$. **f**, Dielectric constant of a 100-nm-thick $PbTiO_3/SrTiO_3$ 2/3 superlattice as a function of temperature. The inset is a close-up of the same data, highlighting the step in the dielectric constant at the transition temperature.

highly desirable from a technological standpoint. In the different context of magneto-electrics and multiferroics, where the magnetic order is frequently associated with rotational structural distortions, similarly changing the nature of the ferroelectric/antiferrodistortive coupling would probably allow tuning of the magnetoelectric response[28–30] in some materials.

METHODS SUMMARY

The theoretical side of this paper required first-principles computational calculations using the ABINIT package to find the unstable modes of the system. Using symmetry considerations, these were related to the experimental results obtained on samples produced using off-axis radio-frequency magnetron sputtering. The phase-transition behaviour in the samples was measured using a laboratory-based X-ray diffractometer and standard electrical characterization equipment.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 August 2007; accepted 30 January 2008.

1. Ahn, C. H., Rabe, K. M. & Triscone, J.-M. Ferroelectricity at the nanoscale: local polarization in oxide thin films and heterostructures. *Science* **303,** 488–491 (2004).
2. Dawber, M., Rabe, K. M. & Scott, J. F. Physics of thin-film ferroelectric oxides. *Rev. Mod. Phys.* **77,** 1083–1130 (2005).
3. Junquera, J. & Ghosez, Ph Critical thickness for ferroelectricity in perovskite ultrathin films. *Nature* **422,** 506–509 (2003).
4. Fong, D. D. *et al.* Ferroelectricity in ultrathin perovskite films. *Science* **304,** 1650–1653 (2004).
5. Lichtensteiger, C., Triscone, J.-M., Junquera, J. & Ghosez, Ph Ferroelectricity and tetragonality in ultrathin $PbTiO_3$ films. *Phys. Rev. Lett.* **94,** 047603 (2005).
6. Rijnders, G. & Blank, D. H. A. Build your own superlattice. *Nature* **433,** 369–370 (2005).
7. Pertsev, N. A., Zembilgotov, A. G. & Tagantsev, A. K. Effect of mechanical boundary conditions on phase diagrams of epitaxial ferroelectric thin films. *Phys. Rev. Lett.* **80,** 1988–1991 (1998).
8. Diéguez, O., Rabe, K. M. & Vanderbilt, D. First-principles study of epitaxial strain in perovskites. *Phys. Rev. B* **72,** 144101 (2005).
9. Choi, K. J. *et al.* Enhancement of ferroelectricity in strained $BaTiO_3$ thin films. *Science* **306,** 1005–1009 (2004).
10. Haeni, J. H. *et al.* Room-temperature ferroelectricity in strained $SrTiO_3$. *Nature* **430,** 758–761 (2004).
11. Neaton, J. & Rabe, K. M. Theory of polarization enhancement in epitaxial $BaTiO_3$/$SrTiO_3$ superlattices. *Appl. Phys. Lett.* **82,** 1586–1588 (2003).
12. Dawber, M. *et al.* Unusual behavior of ferroelectric polarization in $PbTiO_3$/$SrTiO_3$ superlattices. *Phys. Rev. Lett.* **95,** 177601 (2005).
13. Lee, H. N., Christen, H. N., Chrisholm, M. F., Rouleau, C. M. & Lowndes, D. H. Strong polarization enhancement in asymmetric three-component ferroelectric superlattices. *Nature* **433,** 395–399 (2005).
14. Nakhmanson, S. M., Rabe, K. M. & Vanderbilt, D. Polarization enhancement in two- and three-component ferroelectric superlattices. *Appl. Phys. Lett.* **87,** 102906 (2005).
15. Dawber, M. *et al.* Tailoring the properties of artificially layered ferroelectric superlattices. *Adv. Mater.* **19,** 4153–4159 (2007).
16. Kingon, A. I., Maria, J. P. & Streiffer, S. K. Alternative dielectrics to silicon dioxide for memory and logic devices. *Nature* **406,** 1032–1038 (2000).
17. Ghosez, Ph, Cockayne, E., Waghmare, U. V. & Rabe, K. M. Comparative study of the dynamical properties of $BaTiO_3$, $PbTiO_3$ and $PbZrO_3$. *Phys. Rev. B* **60,** 836–843 (1999).
18. Munkholm, A. *et al.* Antiferrodistortive reconstruction of the $PbTiO_3$(001) surface. *Phys. Rev. Lett.* **88,** 016101 (2002).
19. Bungaro, C. & Rabe, K. M. Coexistence of antiferrodistortive and ferroelectric distortions at the $PbTiO_3$ (001) surface. *Phys. Rev. B* **71,** 035420 (2005).
20. Zhong, W. & Vanderbilt, D. Competing structural instabilities in cubic perovskites. *Phys. Rev. Lett.* **74,** 2587–2590 (1995).
21. Pertsev, N. A., Tagantsev, A. K. & Setter, N. Phase transitions and strain-induced ferroelectricity in $SrTiO_3$ epitaxial thin films. *Phys. Rev. B* **61,** R825–R828 (2000).
22. Lin, C.-H., Huang, C.-M. & Guo, G. Y. Systematic *ab initio* study of the phase diagram of epitaxially strained $SrTiO_3$. *J. Appl. Phys.* **100,** 084104 (2006).
23. Vasudevarao, A. *et al.* Multiferroic domain dynamics in strained strontium titanate. *Phys. Rev. Lett.* **97,** 257602 (2006).
24. Gonze, X. *et al.* First-principles computation of material properties: the ABINIT software project. *Computat. Mater. Sci.* **25,** 478–492 (2002).
25. Miller, S. C. & Love, W. F. *Tables of Irreductible Representations of Space Groups and Co-Representations of Magnetic Space Groups* (Pruett, Boulder, 1967).
26. Glazer, A. M. The classification of tilted octahedra in perovskites. *Acta Crystallogr. B* **28,** 3384–3392 (1972).
27. Levanyuk, A. P. & Sannikov, D. G. Improper ferroelectrics. *Uspekhi Fizicheskikh Nauk* **112,** 561–589 (1974).
28. Holakovski, J. A new type of ferroelectric phase transition. *Phys. Status Solidi B* **56,** 615–619 (1973).
29. Eerenstein, W., Mathur, N. D. & Scott, J. F. Multiferroic and magnetoelectric materials. *Nature* **442,** 759–765 (2006).
30. Fennie, C. J. & Rabe, K. M. Ferroelectric transition in $YMnO_3$ from first principles. *Phys. Rev. B* **72,** 100103(R) (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements P.G. thanks A. P. Levanyuk for discussions concerning improper ferroelectrics. We thank R. Černý for help with X-ray diffraction. This work was supported by the VolkswagenStiftung, the European Network of Excellence FAME, the European STREP MaCoMuFi, the Swiss National Science Foundation through the "National Center of Competence in Research Materials with Novel Electronic Properties—MaNEP" and Division II, and ESF(THIOX).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.G. (philippe.ghosez@ulg.ac.be) or M.D. (matthew.dawber@stonybrook.edu).

METHODS

Theory. The first-principles calculations were performed within density functional theory[31] and the local density approximation[31] using the ABINIT package[24]. We used highly transferable Teter pseudopotentials[32]. Strontium 4*s*, 4*p* and 5*s* electrons, Ba 5*s*, 5*p* and 6*s* electrons, Ti 3*s*, 3*p*, 3*d* and 4*s* electrons and O 2*s* and 2*p* electrons were treated as valence states. The wavefunction was expanded in plane waves up to a kinetic energy cutoff of 45 Hartrees. Phonon dispersion curves have been deduced from quantities (dynamical matrix, Born effective charges, optical dielectric constant) obtained using a linear response approach[33]. Differences of polarization were computed using the Berry phase formalism[34]. For the $PbTiO_3/SrTiO_3$ 1/1 superlattice grown along the (001) direction, the primitive unit cell of the prototype *P*4/*mmm* phase contains ten atoms, but for our calculations we used a supercell containing twenty atoms, doubled along [110] to allow the condensation of antiferrodistortive modes. All the calculations concerning the relative stability of the different phases in Fig. 2a of the manuscript were performed within this supercell and integrals over the Brillouin zone were replaced by sums over a $4 \times 4 \times 4$ Monkorst–Pack mesh of special k-points[35], which was checked to be sufficiently accurate.

Experiment. The deposition temperature used for the $PbTiO_3$ and $SrTiO_3$ was approximately 530 °C with an oxygen/argon flow rate of 14 sccm/32 sccm (standard cubic centimetres per minute) and a pressure of 180 mTorr. For the $SrRuO_3$ electrodes the deposition temperature was approximately 560 °C with an oxygen/argon flow rate of 6 sccm/32 sccm and a pressure of 100 mTorr. The polarization and dielectric constant were experimentally measured using an Aixacct TF-Analyser 2000 and an Agilent 4284A LCR meter. The voltage used for switching was 2 V applied across a 100 nm film (the coercive voltage of the films was very low). The data displayed were measured at 1,000 Hz; both the polarization and the dielectric constant were found to be almost independent of measurement frequency in the range 100 Hz to 100 kHz. The X-ray diffraction measurements were carried out on a Panalytical X'Pert Pro MRD diffractometer equipped with an Anton-Parr DHS900 domed hot stage.

31. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge Univ. Press, Cambridge, 2004).
32. Teter, M. P. Additional condition for transferability in pseudopotentials. *Phys. Rev. B* **48,** 5031–5041 (1993).
33. Gonze, X. & Lee, C. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B* **55,** 10355–10368 (1997).
34. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47,** 1651–1654 (1993).
35. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13,** 5188–5192 (1976).

LETTERS

Atmospheric oxidation capacity sustained by a tropical forest

J. Lelieveld[1], T. M. Butler[1], J. N. Crowley[1], T. J. Dillon[1], H. Fischer[1], L. Ganzeveld[1], H. Harder[1], M. G. Lawrence[1], M. Martinez[1], D. Taraborrelli[1] & J. Williams[1]

Terrestrial vegetation, especially tropical rain forest, releases vast quantities of volatile organic compounds (VOCs) to the atmosphere[1–3], which are removed by oxidation reactions and deposition of reaction products[4–6]. The oxidation is mainly initiated by hydroxyl radicals (OH), primarily formed through the photodissociation of ozone[4]. Previously it was thought that, in unpolluted air, biogenic VOCs deplete OH and reduce the atmospheric oxidation capacity[5–10]. Conversely, in polluted air VOC oxidation leads to noxious oxidant build-up by the catalytic action of nitrogen oxides[5–10] ($NO_x = NO + NO_2$). Here we report aircraft measurements of atmospheric trace gases performed over the pristine Amazon forest. Our data reveal unexpectedly high OH concentrations. We propose that natural VOC oxidation, notably of isoprene, recycles OH efficiently in low-NO_x air through reactions of organic peroxy radicals. Computations with an atmospheric chemistry model and the results of laboratory experiments suggest that an OH recycling efficiency of 40–80 per cent in isoprene oxidation may be able to explain the high OH levels we observed in the field. Although further laboratory studies are necessary to explore the chemical mechanism responsible for OH recycling in more detail, our results demonstrate that the biosphere maintains a remarkable balance with the atmospheric environment.

Since forests appeared and proliferated during the Devonian period (approximately 400 million years ago), terrestrial vegetation has been a major contributor to biomass production. At present, terrestrial ecosystems supply more than half the primary production of carbon, dominated by tropical forests and savannahs[11]. The carbon throughput directly affects the atmosphere through the exchange of carbon dioxide, and forests are a major source of volatile organic compounds (VOCs). The main class of compounds comprises the C_5 hydrocarbon isoprene (2-methyl-1,3-butadiene) and polymeric derivates such as C_{10} monoterpenes and C_{15} sesquiterpenes. Annually, plants and trees emit more than a gigatonne of VOC (equivalent to 10^{15} g of carbon), of which isoprene contributes about 40%, several times more than anthropogenic sources[1–3].

The natural VOC emissions serve important biological functions, for example by attracting pollinators and repelling herbivores. The biosphere uses the ambient air as a communication and transport medium, and oxidation of these compounds brings about the concentration gradients sensed by insects. The functionality of isoprene is not fully resolved, although it is known that the molecule sustains the thermal stability of plant cells, counters drought stress and promotes flowering, and that its emission into the atmosphere is strongly dependent on light and temperature[12–15]. After release into the air, isoprene is oxidized within hours by the reaction with OH, the 'detergent' of the atmosphere.

The primary formation of OH results from the photodissociation of ozone by solar ultraviolet radiation in the presence of water vapour[4]. Formation rates are highest in the tropics, where irradiation is intense and the air often humid[6]. After the initial reaction of isoprene with OH, peroxy radicals are formed that either recycle OH by reactions with nitrogen oxide (NO) (pathway I) or recombine into peroxides (pathway II). This is illustrated in Fig. 1, where the peroxides are represented by hydrogen peroxide. The peroxides from pathway II are removed by deposition processes, so that radicals are lost, whereas the recycling through pathway I comes with ozone formation.

However, both pathways have disadvantages for the forest. In the absence of NO in pathway II, the OH radicals and thus the atmospheric oxidation capacity can be depleted. Alternatively, if NO is present in abundance, pathway I generates photochemical smog in which ozone and other toxic oxidants become abundant. Because tropospheric ozone is a powerful greenhouse gas, excess ozone also contributes to climate change. Obviously, the optimum condition for the vegetation is a balance between pathways I and II, but on the basis of the traditional understanding of atmospheric chemistry it is

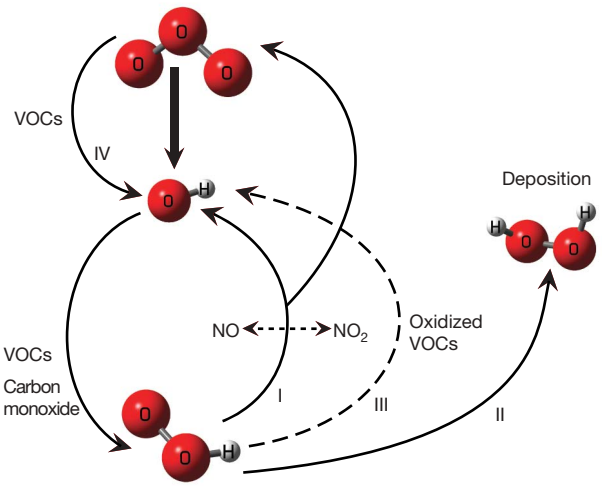


Figure 1 | OH recycling. Photodissociation of ozone leads to primary OH formation (straight arrow). Subsequent OH reactions with carbon monoxide and VOCs produce peroxy radicals. In high-NO conditions OH is recycled and ozone smog builds up through the catalytic conversion between NO and NO_2 (pathway I). In low-NO conditions the deposition of peroxides (pathway II) causes a net loss of OH. We propose that pathway III with oxidized VOCs—in particular organic peroxy radicals—is important in low-NO conditions (dashed arrow). Pathway IV with unsaturated VOCs also occurs, although our results suggest that it has little influence on atmospheric OH.

[1]Max Planck Institute for Chemistry, 27 Becherweg, 55128 Mainz, Germany.

Table 1 | Measured changes in atmospheric chemistry

	Tropical Atlantic		Tropical forest	
	Boundary layer	Free troposphere	Boundary layer	Free troposphere
Isoprene (p.p.b.v.)	0.20 (0.20)	—	2.00 (0.76)	0.07 (0.12)
α-pinene and β-pinene (p.p.b.v.)	—	—	0.42 (0.11)	0.21 (0.08)
CO (p.p.b.v.)	89.2 (10.7)	100.7 (12.8)	113.9 (13.9)	100.8 (21.7)
O_3 (p.p.b.v.)	14.7 (3.3)	33.3 (11.0)	18.5 (4.6)	36.9 (9.9)
NO (p.p.b.v.)	0.01 (0.01)	0.01 (0.01)	0.02 (0.02)	0.02 (0.01)
CH_2O (p.p.b.v.)	0.52 (0.62)	0.34 (0.44)	1.15 (0.86)	0.36 (0.54)
H_2O_2 (p.p.b.v.)	2.43 (0.95)	3.59 (1.09)	2.41 (1.43)	2.76 (1.37)
Organic peroxides (p.p.b.v.)	0.70 (0.27)	0.85 (0.36)	0.96 (0.28)	0.81 (0.25)
CH_3COCH_3 (p.p.b.v.)	0.47 (0.16)	0.54 (0.25)	0.81 (0.24)	0.59 (0.20)
CH_3OH (p.p.b.v.)	1.27 (0.69)	1.82 (0.98)	1.89 (0.78)	1.80 (0.70)
OH (10^6 molecules cm^{-3})	9.0 (4.0)	10.1 (3.1)	5.6 (1.9)	8.2 (3.0)
HO_2 (10^8 molecules cm^{-3})	6.7 (2.3)	5.5 (2.2)	10.5 (2.7)	4.9 (2.0)

The results stated are the mean measurements during daytime, with standard deviations in parentheses. The approximate boundary layer height was 1 km over the ocean and 1.5 km over land. The mixing ratios in the free troposphere are given up to an altitude of 7 km.

difficult to imagine how the atmospheric oxidation capacity could be supported without excess ozone.

Here we propose that the forest makes use of a third, previously overlooked, pathway that combines benefits and circumvents drawbacks (Fig. 1). It appears that in unpolluted low-NO environments isoprene chemistry can directly recycle radicals (pathway III), and the degradation products yield OH more efficiently than was assumed in atmospheric chemistry models. Our hypothesis is based on aircraft measurements of OH radicals and related species in unpolluted air over the Amazon rainforest, laboratory measurements and numerical modelling.

The aircraft measurements were performed in October 2005 between 3–6° N and 50–60° W over the tropical Atlantic Ocean and the pristine forests of Suriname, Guyana and Guyane (French Guiana). Once the coast had been crossed westbound, from the ocean to the forest, the observed boundary layer changes in atmospheric chemistry differed greatly from those expected. The concentration of HO_x ($OH + HO_2$) increased by more than 50%, and that of formaldehyde increased by a factor of two (Table 1). The concentration of hydrogen peroxide remained nearly constant, probably because the higher production rate is compensated by dry deposition, and the concentrations of organic peroxides were enhanced by nearly 40%, probably owing to increasing amounts of peroxides of isoprene. The concentration of methanol increased by about 50%, because of forest emissions, and that of CO by more than 25%, owing to VOC oxidation.

Perhaps most impressively, from about 10 km inland the mean HO_x levels and the diel HO_x cycle were relatively constant, that is, over a distance of at least 1,000 km downwind over the forest, even though local variability can be large. It appears that biosphere–atmosphere interactions quickly reach a steady state, independent of topographical heterogeneity and large biodiversity. Our main finding was that the concentration of OH radicals changed much less than expected. In the tropical marine boundary layer we typically measured 9×10^6 molecules cm^{-3} during daytime, and this decreased by less than 40% over the forest, despite large increases in the concentrations of isoprene and other hydrocarbons. Between the marine and terrestrial free troposphere, the measured differences in concentrations were small.

These measurements are at odds with the results of global atmospheric chemistry models in which the strong biogenic VOC emissions unrealistically deplete OH and isoprene consequently accumulates in the atmospheric boundary layer[5–10]. To circumvent this problem, atmospheric chemistry modellers typically reduce the strength of the isoprene source by a factor of two or more. Such model manipulations hamper the understanding of atmospheric composition changes associated with deforestation and air pollution. Furthermore, it is difficult to imagine how the oxidation capacity was maintained during geological warm epochs when vegetation was copious and VOC emissions probably much stronger than they are today.

To help interpret our results we employed an atmospheric chemistry model with the Mainz isoprene mechanism (MIM)[9,16,17]. Although the primary OH production pathway was accurately reproduced, OH concentrations were strongly underestimated. By testing the MIM against the comprehensive master chemical mechanism[18] we discovered that simplifications commonly applied in atmospheric chemistry models, including the lumping of carbonyls, peroxy radicals and peroxides, caused discrepancies for low-NO environments. By using a new, extended, version of our mechanism,

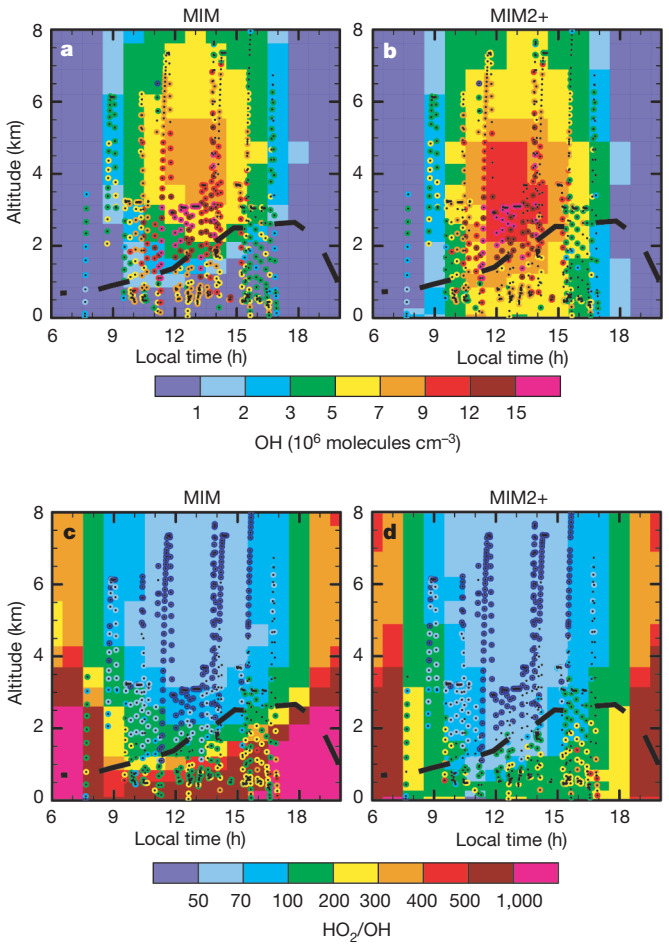


Figure 2 | Model calculations (background) and measurements (circles) of HO_x radicals over Suriname (4° N, 56° W) in October 2005. Daytime OH concentrations (**a**, **b**) and ratios of HO_2 to OH (**c**, **d**). The black dashed lines delineate the upper bound of convective mixing. In **a** and **c** model results are based on the MIM without enhanced OH recycling, and in **b** and **d** on the MIM2+ with 80% OH recycling.

the MIM2, we reached close agreement with the master chemical mechanism, but the total OH formation rate remained too low to explain the observations.

An important improvement was achieved by including recycling of OH by reactions of HO_2 with organic peroxy radicals (RO_2) through pathway III, in accordance with indirect laboratory and field measurements[19–21]. The MIM2 scheme does not include this class of reactions, because they are not present in the master chemical mechanism. We performed laboratory experiments, and by direct detection of OH confirmed a high OH yield of up to ~50% (see Supplementary Information). Although additional experiments will be needed to better characterize isoprene degradation, it seems that several similar reactions may produce OH. Earlier analyses of field data involving isoprene chemistry also pointed to important discrepancies between measured and modelled radical production[21–24].

To test the importance of isoprene-derived peroxy radical reactions we assumed a total OH recycling efficiency of 40–80% in an extended mechanism, the MIM2+. The results in Fig. 2 show that the agreement between the modified model and measurements is good, in contrast with that found using the original MIM (Fig. 2a). Figures 2b and 3b show that the MIM2+ realistically simulates the diel OH cycle. Importantly, the ratio of HO_2 to OH, which is a critical constraint on the chemistry mechanism and greatly overestimated in the boundary layer using the original MIM, was improved dramatically by applying the MIM2+ (Fig. 2c,d). To illustrate the environmental impact, in Fig. 3c we present the global OH changes in the boundary layer.

In addition to isoprene, the forest releases highly reactive terpenes[25]. These contain unsaturated carbon bonds that react with ozone more efficiently than does isoprene and produce OH, shown by pathway IV in Fig. 1. We performed sensitivity studies by implementing terpene–ozone reactions in the model. Although the OH yield of several of these VOCs, such as limonene, myrcene and terpinolene, is close to 100%, their emissions are far too low to increase the concentration of OH in the boundary layer substantially. We also added α-pinene, which is the most abundant monoterpene and reactive towards ozone, to the model. This resulted in a small OH decrease because α-pinene also rapidly reacts with OH.

Although it is conceivable that the forest releases substantial amounts of as-yet-unidentified VOCs[26,27], we believe that their chemical impact is largely limited to the canopy[28] and to the formation of aerosol particles[3]. This is substantiated by the modest concentrations of reaction intermediates such as formaldehyde and organic peroxides in the boundary layer and the presence of a haze layer over the forest. Furthermore, compared to that in the marine boundary layer, the concentration of ozone even increased by a few parts per billion by volume (p.p.b.v.) during transport over the forest (see Table 1), owing to convective mixing with the more ozone-rich free troposphere, which precludes a strong ozone sink by reactions with unsaturated VOCs.

Our research thus points to efficient OH recycling in isoprene chemistry of the order of 40–80%. Nevertheless, additional pathways to OH cannot be ruled out and it seems likely that a combination of factors work in the same direction. For example, it is plausible that multifunctional peroxides, expected products of $RO_2 + HO_2$, photodissociate more efficiently than is currently assumed, because the solar radiation absorption of organic peroxides may be spectrally shifted towards visible irradiances and may thus regenerate OH. This would allow stronger VOC emissions[26,27], without creating a discrepancy between the modelled and measured oxidation products, especially if they condense into aerosol particles.

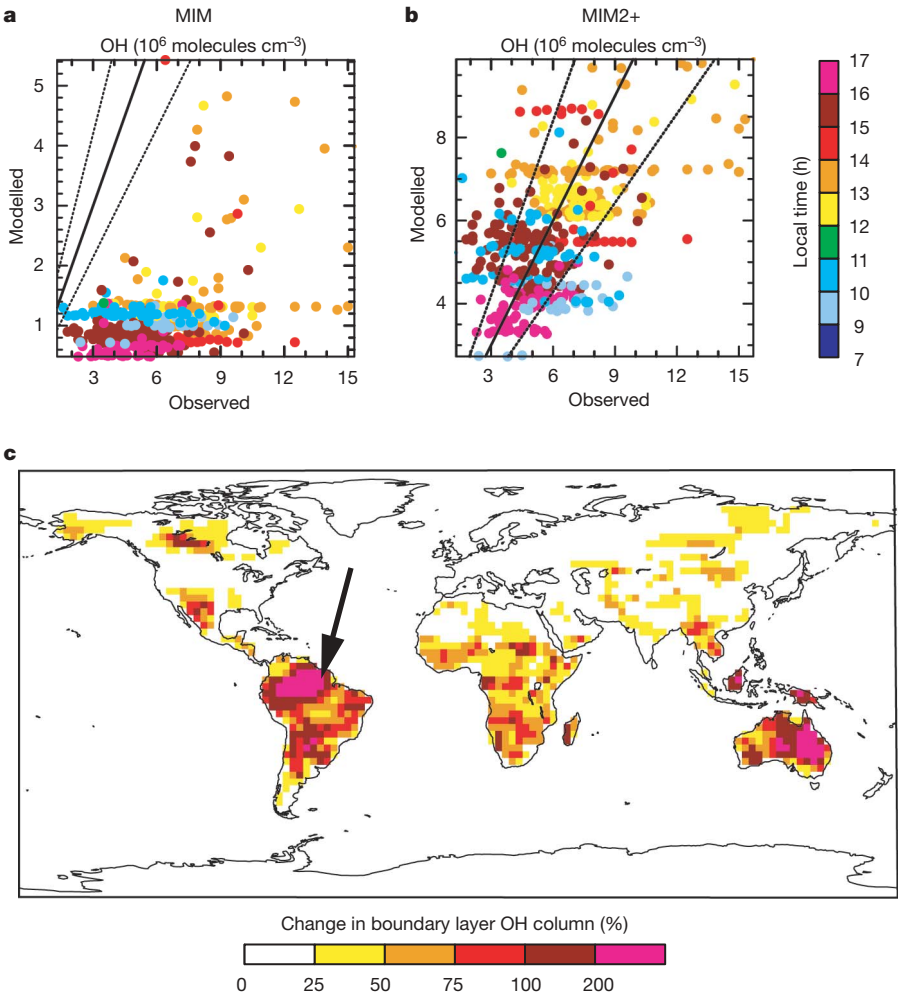


Figure 3 | Difference in OH in the boundary layer, calculated by including enhanced OH recycling in the model. **c**, Percentage difference in the annual mean OH, as calculated using the MIM2+ and the MIM (the arrow indicates the location of Suriname). **a**, **b**, scatter plots between the amounts of OH respectively observed and modelled in the boundary layer over Suriname in October 2005, comparing the MIM (**a**) and the MIM2+ (**b**). The solid lines indicate ideal agreement and the dashed lines the ±40% range, based on the measurement accuracy. Although the global model cannot reproduce the observed small-scale variability over Suriname, the MIM2+ accurately simulates the mean OH and its diel cycle.

We conclude that the biosphere–atmosphere interactions of tropical forest ecosystems maintain a subtle balance. The forest uses the ambient air to transport VOCs for communication and defence, and to dispose of metabolic products while sustaining an auspicious atmospheric environment without the build-up of toxic compounds. Rather than depleting the atmospheric oxidation capacity, as was believed previously, the photochemistry of VOCs over the forest recycles OH in a manner that does not contribute to photochemical smog, in contrast to the anthropogenically influenced atmosphere. Our results show only modest changes in atmospheric oxidation capacity between the marine and terrestrial boundary layers, and differences in the free troposphere are minor. The rapid oxidation by OH of biogenic VOCs that carry sulphur, nitrogen or halogens prevents the loss of nutrients from the ecosystem, because they are locally preserved through the deposition of reaction products. The efficient OH recycling may also explain how the atmospheric oxidation capacity could be sustained in geological warm periods with abundant vegetation.

It is clear that replacing rainforest by agricultural, urban or industrial areas will be associated with enhanced emissions of carbon dioxide, carbon monoxide and NO. In the tropical troposphere the efficient HO_x cycling by NO will strongly enhance photochemical air pollution, posing an additional threat to the remaining biosphere. The resultant excess ozone will also contribute to climate change because ozone is a particularly effective radiative forcing agent in the tropical troposphere. However, in the absence of external influence the forest appears to manage its atmospheric sustainability remarkably well.

METHODS

Field measurement campaign. Our aircraft measurements included OH and HO_2 radicals, NO, ozone, speciated VOCs and oxidation products including ketones, aldehydes and peroxides (see Supplementary Information). The project took place in October 2005, with the operational base in Suriname (5° N, 55° W). There October is relatively dry, with rainfall rates of about 100 mm per month (about half the annual mean), and the regular southeasterly trade winds transport Atlantic air over Guyane, Suriname and Guyana. The region is largely covered by pristine rainforest and local anthropogenic emissions are insignificant, although the free troposphere is occasionally influenced by long-distance transport of biomass-burning emissions. Local perturbations from villages along the coast, small fires or cumulonimbus convection can easily be avoided. This environment was selected to contrast the background marine atmosphere with air that travels over the pristine forest for several days under constant meteorological conditions.

Modelling. We used an atmospheric chemistry and climate model[10] together with a high-resolution single-column model[29] and a chemical box model[17] (see Supplementary Information). Biosphere–atmosphere interactions were computed on-line and the NO and isoprene emission fluxes shown to be realistic for, for example, a rainforest site in the central Amazon[22,29]. Isoprene oxidation reactions in the standard model were described using the MIM[9,16,17], which recycles ~5% of the OH radicals (relative to those consumed in the first isoprene oxidation step). The MIM was first extended to the MIM2, to closely reproduce the master chemical mechanism[18]; the MIM2 recycles ~10% of the OH. An extended version, the MIM2+, was forced to recycle 40–80% by prescribing OH yields from reactions between HO_2 and organic peroxy radicals.

Received 10 October 2007; accepted 22 February 2008.

1. Guenther, A. *et al.* Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmos. Chem. Phys.* **6,** 3181–3210 (2006).
2. Kesselmeier, J. & Staudt, M. Biogenic volatile organic compounds (VOC): An overview on emission, physiology and ecology. *J. Atmos. Chem.* **33,** 23–88 (1999).
3. Goldstein, A. H. & Galbally, I. E. Known and unexplored organic constituents in the Earth's atmosphere. *Environ. Sci. Technol.* **41,** 1515–1521 (2007).
4. Levy, H. II. Normal atmosphere: large radical and formaldehyde concentrations predicted. *Science* **173,** 141–143 (1971).
5. Wang, Y., Jacob, D. J. & Logan, J. A. Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry, 3. Origin of tropospheric ozone and effects of non-methane hydrocarbons. *J. Geophys. Res.* **103,** 10757–10767 (1998).
6. Lelieveld, J., Peters, W., Dentener, F. J. & Krol, M. Stability of tropospheric hydroxyl chemistry. *J. Geophys. Res.* **107,** doi:10.1029/2002JD002272 (2002).
7. Lawrence, M. G. *et al.* A model for studies of tropospheric photochemistry: description, global distribution and evaluation. *J. Geophys. Res.* **104,** 26245–26278 (1999).
8. Granier, C., Petron, G., Müller, J.-F. & Brasseur, G. The impact of natural and anthropogenic hydrocarbons on the tropospheric budget of carbon monoxide. *Atmos. Environ.* **34,** 5255–5270 (2000).
9. von Kuhlmann, R., Lawrence, M. G., Pöschl, U. & Crutzen, P. J. Sensitivities in global scale modelling of isoprene. *Atmos. Chem. Phys.* **4,** 1–17 (2004).
10. Jöckel, P. *et al.* The atmospheric chemistry general circulation model ECHAM5/MESSy: Consistent simulation of ozone from the surface to the mesosphere. *Atmos. Chem. Phys.* **6,** 5067–5104 (2006).
11. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281,** 237–240 (1998).
12. Terry, G. M., Stokes, N. J., Hewitt, C. N. & Mansfield, T. A. Exposure to isoprene promotes flowering in plants. *J. Exp. Bot.* **46,** 1629–1631 (1995).
13. Wildermuth, M. C. & Fall, R. Light-dependent isoprene emission. *Plant Physiol.* **112,** 171–182 (1996).
14. Singsaas, E. L., Lerdau, M., Winter, K. & Sharkey, T. D. Isoprene increases thermotolerance of isoprene emitting species. *Plant Physiol.* **115,** 1413–1420 (1997).
15. Penuelas, J., Lluisa, J., Asensio, D. & Munne-Bosch, S. Linking isoprene with plant thermotolerance, antioxidants and monoterpene emissions. *Plant Cell Environ.* **28,** 278–286 (2005).
16. Pöschl, U., von Kuhlmann, R., Poisson, N. & Crutzen, P. J. Development and intercomparison of condensed isoprene oxidation mechanisms for global atmospheric modelling. *J. Atmos. Chem.* **37,** 29–52 (2000).
17. Sander, R., Kerkweg, A., Jöckel, P. & Lelieveld, J. Technical note: The new comprehensive atmospheric chemistry module MECCA. *Atmos. Chem. Phys.* **5,** 445–450 (2005).
18. Saunders, S. M., Jenkin, M. E., Derwent, R. G. & Pilling, M. J. Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds. *Atmos. Chem. Phys.* **3,** 161–180 (2003).
19. Hasson, A. S., Tyndall, G. S. & Orlando, J. J. A product yield study of the reaction of HO_2 radicals with ethyl peroxy ($C_2H_5O_2$), acetyl peroxy ($CH_3C(O)O_2$), and acetonyl peroxy ($CH_3C(O)CH_2O_2$) radicals. *J. Phys. Chem.* **108,** 5979–5989 (2004).
20. Jenkin, M. E., Hurley, M. D. & Wallington, T. J. Investigation of the radical product channel of the $CH_3COO_2 + HO_2$ reaction in the gas phase. *Phys. Chem. Chem. Phys.* **9,** 3149–3162 (2007).
21. Thornton, J. A. *et al.* Ozone production rates as a function of NO_x abundances and HO_x production rates in the Nashville urban plume. *J. Geophys. Res.* **107,** doi:10.1029/2001JD000932 (2002).
22. Kuhn, U. *et al.* Isoprene and monoterpene fluxes from Central Amazonian rainforest inferred from tower-based and airborne measurements, and implications on the atmospheric chemistry and the carbon budget. *Atmos. Chem. Phys.* **7,** 2855–2879 (2007).
23. Tan, D. *et al.* HO_x budgets in a deciduous forest: Results from the PROPHET summer 1998 campaign. *J. Geophys. Res.* **106,** 24407–24427 (2001).
24. Karl, T. *et al.* The tropical forest and fire emissions experiment: Emission, chemistry and transport of biogenic volatile organic compounds in the lower atmosphere over Amazonia. *J. Geophys. Res.* **112,** doi:10.1029/2007JD008539 (2007).
25. Williams, J., Yassaa, N., Bartenbach, S. & Lelieveld, J. Mirror image hydrocarbons from tropical and boreal forests. *Atmos. Chem. Phys.* **7,** 973–980 (2007).
26. Di Carlo, P. *et al.* Missing OH reactivity in a forest: Evidence for unknown reactive biogenic VOCs. *Science* **304,** 722–725 (2004).
27. Goldstein, A. H. *et al.* Forest thinning experiment confirms ozone deposition to forest canopy is dominated by reaction with biogenic VOCs. *Geophys. Res. Lett.* **31,** doi:10.1029/2004GL021259 (2004).
28. Ciccioli, P. *et al.* Emission of reactive terpene compounds from orange orchards and their removal by within-canopy processes. *J. Geophys. Res.* **104,** 8077–8094 (1999).
29. Ganzeveld, L. N. *et al.* Global soil-biogenic NO_x emissions and the role of canopy processes. *J. Geophys. Res.* **107,** doi:10.1029/2001JD001289 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the GABRIEL campaign team: S. Bartenbach, C. Becker, H. Bozem, S. Engemann, H. Franke, S. Gebhardt, C. Gurk, H. Hoeseni, R. Hofmann, T. Klüpfel, R. Königstedt, D. Kubistin, R. Maser, D. Noorden, U. Parchatka, D. Rodrigues, M. Rudolf, B. Scheeren, C. Schiller, V. Sinha, A. Stickler, B. Tan, P. van Velthoven, T. Warsodikromo and G. Wesenhagen. We thank the Modular Earth Submodel System (MESSy) team for model support, in particular P. Jöckel and H. Tost, and P. J. Crutzen for comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.L. (lelieveld@mpch-mainz.mpg.de).

LETTERS

SAR11 marine bacteria require exogenous reduced sulphur for growth

H. James Tripp[1], Joshua B. Kitner[1], Michael S. Schwalbach[1], John W. H. Dacey[2], Larry J. Wilhelm[1] & Stephen J. Giovannoni[1]

Sulphur is a universally required cell nutrient found in two amino acids and other small organic molecules. All aerobic marine bacteria are known to use assimilatory sulphate reduction to supply sulphur for biosynthesis, although many can assimilate sulphur from organic compounds that contain reduced sulphur atoms[1–3]. An analysis of three complete '*Candidatus* Pelagibacter ubique' genomes, and public ocean metagenomic data sets, suggested that members of the ubiquitous and abundant SAR11 alphaproteobacterial clade are deficient in assimilatory sulphate reduction genes. Here we show that SAR11 requires exogenous sources of reduced sulphur, such as methionine or 3-dimethylsulphoniopropionate (DMSP) for growth. Titrations of the algal osmolyte DMSP in seawater medium containing all other macronutrients in excess showed that 1.5×10^8 SAR11 cells are produced per nanomole of DMSP. Although it has been shown that other marine alphaproteobacteria use sulphur from DMSP in preference to sulphate[1,2], our results indicate that '*Cand.* P. ubique' relies exclusively on reduced sulphur compounds that originate from other plankton.

SAR11 is the most abundant and ubiquitous clade of heterotrophic marine bacteria in the oceans, often comprising 30% of the surface bacterial plankton community[4]. '*Cand.* P. ubique' strain HTCC1062, as well as other cultivated strains of SAR11, consistently grow to about 1.12×10^6 cells ml^{-1} ($n = 41$, s.d. $\pm 4.04 \times 10^5$, range 2.5×10^5 to 3.5×10^6 cells ml^{-1}) on natural, autoclaved sea water amended with ammonium, phosphate and various organic carbon compounds, under dark or light conditions. These cell densities are similar to SAR11 population densities observed in coastal Oregon sea water, suggesting that, when growing in seawater culture, SAR11 strains are limited by compounds found naturally in the environment. Iron is also thought to sometimes limit biomass production in marine environments[5], but the addition of 53.6 μM $FeCl_3$ did not affect '*Cand.* P. ubique' yields (1.22×10^6 cells ml^{-1}, $n = 2$, s.d. $\pm 3.89 \times 10^5$), nor did additions of 18 trace elements used in culturing marine bacteria[6] (data not shown).

The reconstruction of metabolic pathways from complete genome sequences of two '*Cand.* P. ubique' strains, HTCC1002 and HTCC1062, revealed an apparently incomplete set of genes for assimilatory sulphate reduction in both genomes, suggesting that SAR11 might require reduced sulphur compounds for growth. The canonical assimilatory reduction pathway for *Escherichia coli* shown in Fig. 1a consists of *cysDNCHIJ*, which occurs in two operons (Fig. 1b). Variants of the canonical pathway exist. *Corynebacterium*

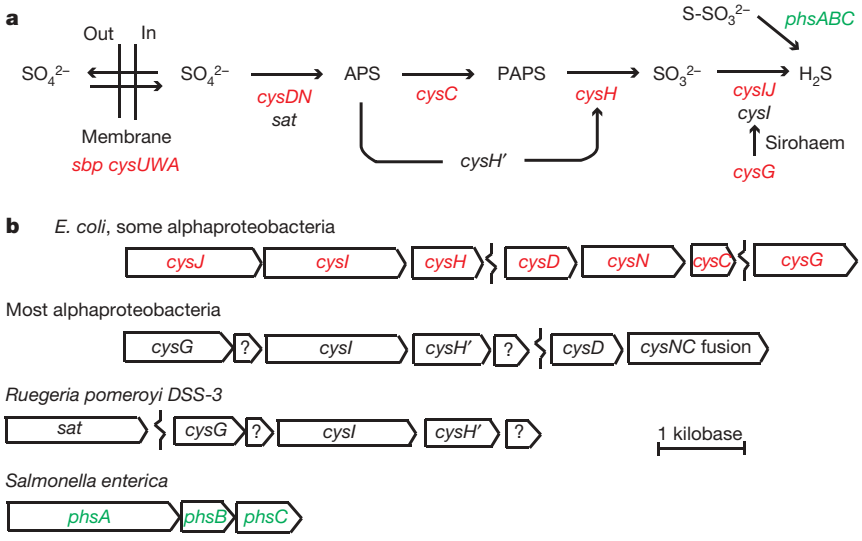


Figure 1 | Sulphur metabolism pathways and comparative genomics. **a**, Pathways of sulphate transport, assimilatory sulphate reduction, thiosulphate reduction and sulphite oxidation. *E. coli* genes for canonical assimilatory sulphate reduction and sulphate transport are shown in red, alternative assimilatory sulphate reduction genes are shown in black, *Salmonella* genes for canonical thiosulphate reduction are shown in green. APS, adenosine 5′-phosphosulphate; PAPS, 3′-phosphoadenosine 5′-phosphosulphate. **b**, Operon organization of assimilatory sulphate reduction pathway from selected organisms.

[1]Department of Microbiology, 220 Nash Hall, Oregon State University, Corvallis, Oregon 97331, USA. [2]Woods Hole Oceanographic Institution, Redfield 3-22, MS no. 32, Woods Hole, Massachusetts 02543, USA.

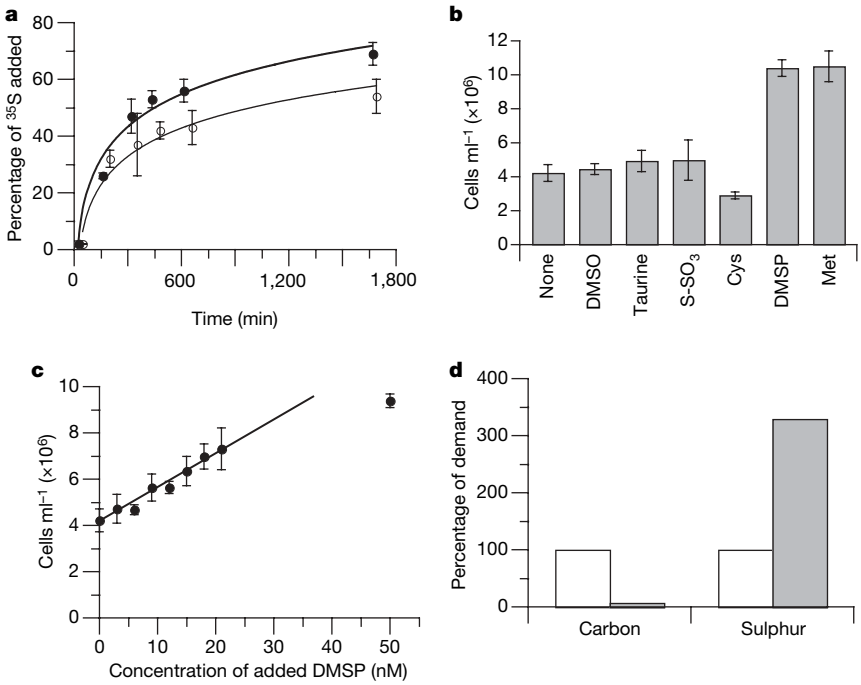


Figure 2 | ^{35}S radiotracer uptake and growth responses to additions of sulphur compounds. **a**, Incorporation of ^{35}S from ^{35}S[DMSP] into protein. Filled symbols, ^{35}S uptake ($R^2 = 0.9734$); open symbols, ^{35}S incorporation into trichloroacetic acid-insoluble fraction ($R^2 = 0.9586$). **b**, Maximum cell density for additions of 100 nM sulphur compounds. Control is sea water and excess carbon, nitrogen, phosphorus, iron and vitamins. The sulphur compounds are shown in order from most oxidized to most reduced. Cysteine and thiosulphate results varied depending on the batch of sea water, with cysteine showing a slight inhibition in this experiment. **c**, Molar growth yield for DMSP with other nutrients in excess. The regression equation is $y = 150{,}797x + (4 \times 10^6)$; $R^2 = 0.9683$. **d**, Calculated carbon and sulphur supply from 1 nM DMSP enrichment compared with the biomass demand, as described in the text. Error bars in **a**–**c** show s.d. for triplicate samples.

and *Allochromatium* do not seem to need *cysJ*[7,8], nor apparently do most alphaproteobacteria, on the basis of alignments viewed in MicrobesOnline[9]. *Bacillus subtilis* lacks the adenylylsulphate kinase gene *cysC*[10], and uses '*sat*' genes instead of *cysDN*, an arrangement also observed in the marine alphaproteobacterium *Ruegeria* (formerly *Silicibacter*) *pomeroyi* DSS-3. *Salmonella enterica* can reduce thiosulphate to sulphide under anaerobic conditions, using *phsABC*[11]. A tBLASTx search of relevant nucleotide sequences from *Salmonella*, *Ruegeria*, *Caulobacter crescentus* and *E. coli* against HTCC1062 and HTCC1002 revealed no putative genes for assimilatory sulphate reduction (see Supplementary Information). On the basis of this evidence, we predicted that SAR11 was deficient in assimilatory sulphate reduction.

Because multiple transporters and putative degradation pathways were found for DMSP and methionine in the SAR11 genome (see Supplementary Information), we proposed that DMSP or methionine might serve as reduced sulphur sources, and designed experiments to test these hypotheses. DMSP uptake in natural SAR11 populations has been established previously by microradiography[12], and DMSP demethylation activity has been demonstrated in cloned SAR11 genes[13]. Bacteria possessing the gene for DMSP demethylation (*dmdA*) demethylate DMSP and use the remaining methanethiol group for methionine biosynthesis[1]. Bacteria possessing the marker gene for the DMSP lyase pathway (*dddD*) cleave DMSP to form 3-hydroxypropionate and dimethylsulphide (DMS), a volatile gas that oxidizes to products that cause clouds to form[14]. It is estimated that 50–90% of the DMSP metabolized by marine microorganisms is channelled into the demethylation pathway, thereby mitigating the impact of DMSP on climate[13]. On a global basis, the SAR11 clade is likely to be responsible for a significant fraction of the uptake and demethylation of DMSP because of its abundance and ubiquity in the oceans.

'*Cand.* P. ubique' cells readily incorporated ^{35}S from ^{35}S[DMSP] into cellular protein (Fig. 2a). The percentages of the added ^{35}S that were taken up (70%) and incorporated (50%) are similar to the highest values reported previously[15] (63% and 55%) for subnanomolar additions of DMSP to natural plankton communities.

Additions of 100 nM DMSP or methionine, with all other nutrients in excess, resulted in cell densities of just over 10^7 cells ml^{-1} ($n = 3$, shown in Fig. 2b), indicating that these compounds could relieve nutrient limitation. The response to additions of different concentrations of cysteine and thiosulphate to media collected on different dates was variable, as a result of factors that are not understood. The addition of cysteine sometimes increased yield (Fig. 3), but in other experiments with different batches of sea water it did not (Fig. 2b). A pathway for the conversion of cysteine to methionine

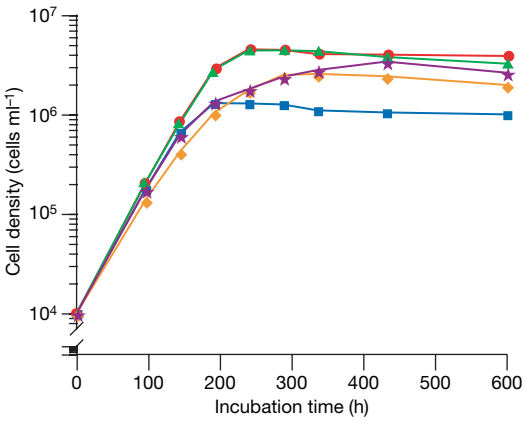


Figure 3 | Growth curves for 21-nM additions of various reduced sulphur sources. Data points show averages of duplicate measurements from two treatments. The control was sea water to which an excess of defined organic carbon compounds, nitrogen, phosphorus, iron and vitamins had been added. Blue, control; orange, cysteine; purple, thiosulphate; green, methionine; red, DMSP.

Table 1 | Other aerobes that may be deficient in assimilatory sulphate reduction

Organism	Phylum	Class	Genome size* (kbp)	Habitat	Phenotype
Halobacterium NRC-1	Euryarchaeota	Halobacteria	2,571	Ponds, aquatic	Aerobe, chemo-organotroph, motile, halophile, rod-shaped
Picrophilus torridus	Euryarchaeota	Thermoplasma	1,545	Soil	Acidophile, thermophile, aerobe, heterotroph
Idiomarina iloihiensis	Proteobacteria	Gamma	2,839	Aquatic, hydrothermal vent	Halophile, rod-shaped, aerobe

* From Genomes OnLine Database (GOLD)[30]. kbp, kilobase pairs.

could not be confirmed from the genome annotation of '*Cand.* P. ubique' strains HTCC1062 and HTCC1002, suggesting that cysteine might not be able to supply the cell's full requirement for reduced sulphur (see Supplementary Information). Regardless of sulphur source (DMSP, methionine, cysteine or thiosulphate), the growth rate during the exponential phase ranged from 0.62 to 0.74 d^{-1}, in comparison with 0.70 d^{-1} for the control (Fig. 3). Extension of the exponential phase resulted in increased yields ranging from 2.56×10^6 cells ml^{-1} to 4.52×10^6 cells ml^{-1}, in contrast with 1.3×10^6 cells ml^{-1} for the control.

The yield of SAR11 cells as a function of DMSP concentration in the presence of vitamins, excess organic carbon and excess inorganic nitrogen, phosphorus and iron is reported in Fig. 2c. As shown by the slope of the linear regression, the maximum cell density increased by 1.5×10^5 cells ml^{-1} per 1 nM increase in DMSP concentration. Measurements of DMSP in the autoclaved seawater medium ranged from 0.3 to 0.7 nM, indicating that DMSP can support population sizes of 4.5×10^4 to 1.05×10^5 cells ml^{-1}. We therefore infer that SAR11 cells cultured in autoclaved sea water are likely to be using sulphur compounds in addition to DMSP to supply their sulphur requirement.

To ensure that sulphur, not carbon, from DMSP was responsible for the increase in biomass, we estimated biomass demands from a recent measurement of the '*Cand.* P. ubique' cell volume obtained by cryoelectron tomography[16]. Assuming a density of 1 g cm^{-3}, carbon as half the dry weight and a sulphur/carbon ratio of 1:90 (ref. 2), the carbon mass for a cell of 0.035 μm^3 is 5.8 fg and the sulphur mass per cell is 65 ag (65×10^{-18} g). Each 1 nM increase in DMSP concentration yielded an additional 1.5×10^5 cells ml^{-1}, or an additional 1.5×10^8 cells l^{-1}. Although the medium contained an excess of added organic carbon, we estimated carbon and sulphur demand per unit of cell biomass compared with the supply of these elements provided by the added DMSP. DMSP sulphur was more than sufficient (329%) to meet the estimated sulphur demand, but the carbon supplied by the added DMSP fell well short of the estimated carbon demand for biomass (7%), as shown in Fig. 2d. Therefore the carbon in DMSP could not have been solely responsible for the observed increase in cell yield. Excess consumption of sulphur relative to the calculated demand has been reported from studies of natural populations of marine bacteria[15,17,18] and may be due to variances in carbon-to-sulphur ratios or methods of measurement. Cell morphology and size were consistent in all media and stages of growth, up to the maximum cell density (see Supplementary Information).

The SAR11 clade is a diverse group, raising the question of whether a nutritional requirement for organic compounds containing reduced sulphur is a common feature of this clade. To address this, we searched metagenomic DNA from the Global Ocean Survey for the *cysDNCHIJ*, *sat* and *phsABC* nucleotide sequences mentioned above by using tBLASTx and a bit score cutoff of 200. This comprehensive list of all the fragments in the metagenome containing assimilatory sulphate reduction genes was then searched for any gene with a best hit to an HTCC1062 gene. Because many of the random DNA fragments in metagenomic data encode regions of two or more genes, in principle this procedure should detect instances of the insertion of *cysDNCHIJ*, *sat* and *phsABC* into the context of SAR11 genes known from the available complete genome sequences. Only one fragment, which contained a *Caulobacter cysD* gene, was found adjacent to a gene having a best hit to an HTCC1062 gene. Additional support for the conclusion that assimilatory sulphate reduction is deficient in many SAR11 populations comes from the genome sequence of a newly isolated Sargasso Sea strain of SAR11 that also has no conserved domain sequences with significant BLAST hits to canonical assimilatory sulphate reduction genes (U. Stingl, personal communication). These data strongly suggest a requirement for exogenous reduced organic sulphur that is widespread, if not universal, within the SAR11 clade.

The absence of assimilatory sulphate reduction has been reported in anaerobic photoautotrophs from the green sulphur[19], purple nonsulphur[20] and purple sulphur[21] bacteria, and in methanogenic archaea[22], all of which live in sulphide-rich environments. The loss of assimilatory sulphate reduction in pathogens was shown convincingly in the endosymbiont *Buchnera aphidicola*[23]. We searched complete genome sequences for other free-living aerobes that might also be deficient for assimilatory sulphate reduction and found two archaea and one bacterium that were missing essential genes (Table 1). Hou *et al.*[24] could not find canonical assimilatory sulphate reduction genes in the free-living aerobic marine gammaproteobacterium *Idiomarina iloihiensis*, and listed four small subunits of dissimilatory sulphate reduction genes as candidates for performing this function.

We previously reported that SAR11 has the smallest genome known for a free-living heterotrophic cell and has a low proportion of non-coding DNA[25]. These observations were attributed to the genome-streamlining hypothesis, which postulates that in very large microbial populations in which nitrogen and phosphorus are often limiting resources, selection efficiently eliminates DNA that has no function. Here we show that the absence of a conserved metabolic pathway that is not essential for autonomous replication in sea water is also a factor that contributes to the small genome size of '*Cand.* P. ubique'. Assimilatory sulphate reduction is nearly universal in aerobic marine bacteria, probably because sulphate is found at saturating concentrations in sea water, and the assimilatory sulphate reduction pathway ensures that cells will have a supply of sulphur for biosynthesis. However, the double negative charge on sulphate makes it more expensive to import, and it requires eight electrons for reduction to sulphide. Presumably, during the evolution of the SAR11 clade the loss of nutritional autonomy associated with the elimination of assimilatory sulphate reduction was offset by the advantages of a smaller genome and a metabolic strategy that acquires sulphur at a lower bioenergetic cost.

The benefits of genome reduction are probably not without a cost. The elimination of assimilatory sulphate reduction genes from SAR11 makes them dependent on reduced sulphur compounds found in the environment. One implication of our findings is that natural SAR11 populations may undergo transient periods where their growth is limited by the availability of sources of reduced sulphur, placing them at a competitive disadvantage alongside organisms that have the assimilatory sulphate reduction pathway. Understanding the factors that control plankton populations is a grail for oceanographers. It seems likely that the findings we report will focus attention on reduced sulphur compounds as potentially significant factors influencing microbial population dynamics in the oceans.

METHODS SUMMARY

Culturing and counting. Culturing techniques and seawater collection are described elsewhere[26,27]. The mixture of defined organic carbon compounds added to the medium to ensure an excess of carbon was the following: 100 nM

glycine betaine; 56 μM glucose, 44 μM ribose, 85 μM succinate, 114 μM pyruvate, 109 μM glycerol and 45 μM *N*-acetyl glucosamine; and 434 μM ethanol. Excess nitrogen and phosphorus were given as 10 μM NH_4Cl and 1 μM KH_2PO_4. Excess iron was given as 53.6 μM $FeCl_3$. Cell counts were acquired on an EasyCyte flow cytometer (Guava Technologies) after 1 h of staining in 1:2,000 diluted SYBR-Green I (Invitrogen)[28]. Sulphur compound comparisons were made at 100 nM in the same water, collected from the Oregon coast on 15 March 2007, as that used for the molar growth yield curve, which supported unusually high SAR11 cell densities in cultures. Experiments with other water samples gave similar results. Cells were incubated for three weeks to obtain the molar growth yield curve and maximum cell densities for various sulphur sources.

Radiotracer uptake. ^{35}S[DMSP] (10 μl; specific radioactivity 380,684 d.p.m. pmol^{-1}, final concentration 1.8 pM) was added to triplicate 600-ml cultures containing mid-exponential cells growing at 3.3×10^5 cells ml^{-1}. The cultures were incubated at 20 ± 0.5 °C for the duration of the experiment. For the isotope uptake measurements, duplicate 10-ml samples were filtered on 0.2-μm polycarbonate filters, rinsed twice with 10 ml of sea water, and measured by liquid-scintillation counting. To measure the incorporation of the isotope into protein, filters were also washed twice with 5 ml of ice-cold trichloroacetic acid before being counted. For each time point, cells were stained with 4,6-diamidino-2-phenylindole and counted by epifluorescence microscopy. The final cell density after 28 h was 7.6×10^5 cells ml^{-1}.

Search for candidate genomes. We began with a Genome Properties query for GenProp0149 'sulphate reduction to sulphide, assimilatory' in the Comprehensive Microbial Resource (CMR)[29]. Of the 381 genomes listed, 147 showed 'none found' for the genes present. We eliminated genomes with 'pathogen' or 'anaerobic' in the 'phenotype' field of the Genomes OnLine Database (GOLD)[30] and used the MicrobesOnline[9] homologue finder and genome browser to eliminate genomes that had assimilatory sulphate reduction operons apparently missed by CMR.

Received 19 June 2007; accepted 30 January 2008.
Published online 12 March 2008.

1. Kiene, R. P., Linn, L. J., Gonzalez, J., Moran, M. A. & Bruton, J. A. Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. *Appl. Environ. Microbiol.* **65,** 4549–4558 (1999).
2. Cuhel, R. L., Taylor, C. D. & Jannasch, H. W. Assimilatory sulfur metabolism in marine microorganisms: sulfur metabolism, protein synthesis, and growth of *Alteromonas luteo-violaceus* and *Pseudomonas halodurans* during perturbed batch growth. *Appl. Environ. Microbiol.* **43,** 151–159 (1982).
3. Cottrell, M. T. & Kirchman, D. L. Natural assemblages of marine proteobacteria and members of the Cytophaga–Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.* **66,** 1692–1697 (2000).
4. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420,** 806–810 (2002).
5. Street, J. H. & Paytan, A. Iron, phytoplankton growth, and the carbon cycle. *Met. Ions Biol. Syst.* **43,** 153–193 (2005).
6. Schut, F. *et al.* Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. *Appl. Environ. Microbiol.* **59,** 2150–2160 (1993).
7. Neumann, S., Wynen, A., Truper, H. G. & Dahl, C. Characterization of the *cys* gene locus from *Allochromatium vinosum* indicates an unusual sulfate assimilation pathway. *Mol. Biol. Rep.* **27,** 27–33 (2000).
8. Ruckert, C. *et al.* Functional genomics and expression analysis of the *Corynebacterium glutamicum* fpr2-cysIXHDNYZ gene cluster involved in assimilatory sulphate reduction. *BMC Genomics* **6,** 121 (2005).
9. Alm, E. J. *et al.* The MicrobesOnline Web site for comparative genomics. *Genome Res.* **15,** 1015–1022 (2005).
10. Berndt, C. *et al.* Characterization and reconstitution of a 4Fe-4S adenylyl sulfate/phosphoadenylyl sulfate reductase from *Bacillus subtilis*. *J. Biol. Chem.* **279,** 7850–7855 (2004).
11. Heinzinger, N. K., Fujimoto, S. Y., Clark, M. A., Moreno, M. S. & Barrett, E. L. Sequence analysis of the *phs* operon in *Salmonella typhimurium* and the contribution of thiosulfate reduction to anaerobic energy metabolism. *J. Bacteriol.* **177,** 2813–2820 (1995).
12. Malmstrom, R. R., Kiene, R. P., Cottrell, M. T. & Kirchman, D. L. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. *Appl. Environ. Microbiol.* **70,** 4129–4135 (2004).
13. Howard, E. C. *et al.* Bacterial taxa that limit sulfur flux from the ocean. *Science* **314,** 649–652 (2006).
14. Todd, J. D. *et al.* Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **315,** 666–669 (2007).
15. Kiene, R. P. & Linn, L. J. The fate of dissolved dimethylsulfoniopropionate (DMSP) in seawater: Tracer studies using S-35-DMSP. *Geochim. Cosmochim. Acta* **64,** 2797–2810 (2000).
16. Nicastro, D. *et al.* Three-dimensional structure of the tiny bacterium *Pelagibacter ubique* studied by cryo-electron tomography. *Microsc. Microanal.* **12,** 180–181 (2006).
17. Simo, R., Archer, S. D., Pedros-Alio, C., Gilpin, L. & Stelfox-Widdicombe, C. E. Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. *Limnol. Oceanogr.* **47,** 53–61 (2002).
18. Zubkov, M. V. *et al.* Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. *Deep-Sea Res. II* **49,** 3017–3038 (2002).
19. Lippert, K. D. & Pfennig, N. Utilisation of molecular hydrogen by *Chlorobium thiosulfatophilum*. Growth and CO_2-fixation. *Arch. Mikrobiol.* **65,** 29–47 (1969).
20. Glaeser, J. & Overmann, J. Selective enrichment and characterization of *Roseospirillum parvum*, gen. nov. and sp. nov., a new purple nonsulfur bacterium with unusual light absorption properties. *Arch. Microbiol.* **171,** 405–416 (1999).
21. Ventura, S., De Philippis, R., Materassi, R. & Balloni, W. Two halophilic *Ectothiorhodospira* strains with unusual morphological, physiological and biochemical characters. *Arch. Microbiol.* **149,** 273–279 (1988).
22. Daniels, L., Belay, N. & Rajagopal, B. S. Assimilatory reduction of sulfate and sulfite by methanogenic bacteria. *Appl. Environ. Microbiol.* **51,** 703–709 (1986).
23. van Ham, R. C. *et al.* Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl Acad. Sci. USA* **100,** 581–586 (2003).
24. Hou, S. *et al.* Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl Acad. Sci. USA* **101,** 18036–18041 (2004).
25. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309,** 1242–1245 (2005).
26. Rappe, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418,** 630–633 (2002).
27. Connon, S. A. & Giovannoni, S. J. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* **68,** 3878–3885 (2002).
28. Stingl, U., Tripp, H. J. & Giovannoni, S. J. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J* **1,** 361–371 (2007).
29. Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K. & White, O. The comprehensive microbial resource. *Nucleic Acids Res.* **29,** 123–125 (2001).
30. Bernal, A., Ear, U. & Kyrpides, N. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res.* **29,** 126–127 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Stingl, C. Carlson and A. Treusch for discussions, R. Kiene for ^{35}S[DMSP], and K. Vergin for assistance with radiotracers. This work was supported by grants from the Marine Microbiology Initiative of the Gordon and Betty Moore Foundation and the National Science Foundation.

Author Contributions H.J.T. conducted genome analysis, adaptation of Guava EasyCyte to marine picoplankton counting, initial experimental design, preliminary data collection and analysis, and wrote the manuscript. J.B.K. set up, performed, and collected data for many culture experiments, and assisted in EasyCyte adaptations. J.W.H.D. provided DMSP and DMSP measurements. L.J.W. performed metagenomics. S.J.G. initiated the study, suggested candidate nutrients for testing, and with M.S.S. suggested final experiments and data presentation. All authors reviewed the manuscript before submission.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.J.G. (steve.giovannoni@oregonstate.edu).

LETTERS

Broad phylogenomic sampling improves resolution of the animal tree of life

Casey W. Dunn[1]†, Andreas Hejnol[1], David Q. Matus[1], Kevin Pang[1], William E. Browne[1], Stephen A. Smith[2], Elaine Seaver[1], Greg W. Rouse[3], Matthias Obst[4], Gregory D. Edgecombe[5], Martin V. Sørensen[6], Steven H. D. Haddock[7], Andreas Schmidt-Rhaesa[8], Akiko Okusu[9], Reinhardt Møbjerg Kristensen[10], Ward C. Wheeler[11], Mark Q. Martindale[1] & Gonzalo Giribet[12,13]

Long-held ideas regarding the evolutionary relationships among animals have recently been upended by sometimes controversial hypotheses based largely on insights from molecular data[1,2]. These new hypotheses include a clade of moulting animals (Ecdysozoa)[3] and the close relationship of the lophophorates to molluscs and annelids (Lophotrochozoa)[4]. Many relationships remain disputed, including those that are required to polarize key features of character evolution, and support for deep nodes is often low. Phylogenomic approaches, which use data from many genes, have shown promise for resolving deep animal relationships, but are hindered by a lack of data from many important groups. Here we report a total of 39.9 Mb of expressed sequence tags from 29 animals belonging to 21 phyla, including 11 phyla previously lacking genomic or expressed-sequence-tag data. Analysed in combination with existing sequences, our data reinforce several previously identified clades that split deeply in the animal tree (including Protostomia, Ecdysozoa and Lophotrochozoa), unambiguously resolve multiple long-standing issues for which there was strong conflicting support in earlier studies with less data (such as velvet worms rather than tardigrades as the sister group of arthropods[5]), and provide molecular support for the monophyly of molluscs, a group long recognized by morphologists. In addition, we find strong support for several new hypotheses. These include a clade that unites annelids (including sipunculans and echiurans) with nemerteans, phoronids and brachiopods, molluscs as sister to that assemblage, and the placement of ctenophores as the earliest diverging extant multicellular animals. A single origin of spiral cleavage (with subsequent losses) is inferred from well-supported nodes. Many relationships between a stable subset of taxa find strong support, and a diminishing number of lineages remain recalcitrant to placement on the tree.

Expressed sequence tags (ESTs) provide opportunities to sample diverse genes from a large number of taxa[6]. Several recent phylogenomic studies, based largely on EST data, analysed matrices containing more than 140 genes from up to 34 metazoans (multicellular animals)[7–9]. However, the included species were not well sampled across extant metazoan diversity. These analyses also relied on either ribosomal proteins or a list of target genes identified from a small (1,152 ESTs) choanoflagellate data set[10], limiting the possibilities of EST studies to inform gene selection and homology assignment. Rather than look for predefined sets of genes in our data, we present an explicit procedure for gene selection (see Methods and Supplementary Fig. 2).

Our complete matrix includes data from 77 taxa (of which 71 are metazoans) and 150 genes. On average, taxa in our matrix include 50.9% of the 150 genes, and overall matrix completeness is 44.5%. Maximum likelihood (WAG model of sequence evolution; Figs 1 and 2) and bayesian (CAT[11] and WAG models of sequence evolution; Fig. 2) analyses of our matrix support the major groups of the 'new animal phylogeny'[2]. These groups have also been supported by other EST-based analyses[9], but not by phylogenomic studies that consider a small number of animal taxa[12]. Primary analyses of the 77-taxon matrix recover Metazoa, Bilateria and Protostomia with strong bootstrap support (>90%). This is an improvement compared to some previous phylogenomic studies that did not recover Protostomia, which in part led one study to conclude that it may not be possible to reconstruct the relationships of several major clades of animals because the metazoan radiation was too rapid[13]. It now seems that those findings were largely caused by limited taxon sampling, a result consistent with reanalyses[14]. Bootstrap support for Lophotrochozoa and Ecdysozoa is low in the 77-taxon consensus tree, but this is caused by the instability of a relatively small number of taxa (see below). Whereas Deuterostomia had poor support in recent phylogenomic analyses[15], in analyses of our 77-taxon matrix maximum likelihood bootstrap support for Deuterostomia is >80%. Within Deuterostomia, *Xenoturbella* was found to be sister to Ambulacraria (echinoderms and hemichordates) in a study that included 1,372 *Xenoturbella* ESTs[7]. Our inclusion of 3,840 additional *Xenoturbella* ESTs is consistent with this previous analysis (Figs 1, 2). None of our results are congruent with Coelomata, a group consisting of taxa that have a coelomic body cavity, which was favoured before molecular data became available. Coelomata has been recovered in some studies using many genes from a very small number of taxa[12,16], but it now seems clear that this is an artefact of poor taxon sampling.

Low-support values on consensus trees can be caused by large-scale structural rearrangements or by the instability of particular taxa. If, for instance, a taxon is only placed within a particular clade 50% of

[1]Kewalo Marine Laboratory, PBRC, University of Hawaii, 41 Ahui Street, Honolulu, Hawaii 96813, USA. [2]Department of Ecology and Evolutionary Biology, Yale University, PO Box 208105, New Haven, Connecticut 06520, USA. [3]Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive 0202, La Jolla, California 92093, USA. [4]Kristineberg Marine Research Station, Kristineberg 566, 450 34 Fiskebäckskil, Sweden. [5]Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. [6]Ancient DNA and Evolution Group, Biological Institute, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. [7]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, California 95039, USA. [8]Zoological Museum, University of Hamburg, Martin-Luther-King-Platz 3, 20146 Hamburg, Germany. [9]Biology Department, Simmons College, The Fenway, Boston, Massachusetts 02115, USA. [10]Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. [11]Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA. [12]Department of Organismic and Evolutionary Biology, [13]Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA. †Present address: Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman Street, Providence, Rhode Island 02912, USA.

the time, the support for that clade will be 50%, even if all other features of the tree are identical. This can obscure strongly supported relationships among stable taxa. We therefore used quantitative criteria to remove unstable taxa by calculating leaf stability indices[17], which measure the consistency of a taxon's position relative to other taxa across replicates, for all ingroup taxa (Fig. 1) and generated a new 64-taxon data set including only the most stable taxa (leaf stability, >90%). Some of the 13 unstable taxa (Entoprocta, Myzostomida, the sponge *Suberites domuncula* and the acoels) had poor gene sampling (Supplementary Tables 1 and 2, and Supplementary Fig. 3), which may simply provide too few informative characters for phylogenetic reconstruction. Acoels have also been found to be unstable in other phylogenomic studies[15]. Other unstable taxa (for example, Rotifera, Bryozoa and Gnathostomulida) had good gene sampling, suggesting that improved taxon sampling may be the most promising strategy for resolving their positions. Most unstable taxa moved between only a few positions (Supplementary Fig. 8), with most placed closer to Platyhelminthes than to other stable taxa, recovering with poor support a group known as Platyzoa[18]. Platyhelminths have relatively long branches, and it may

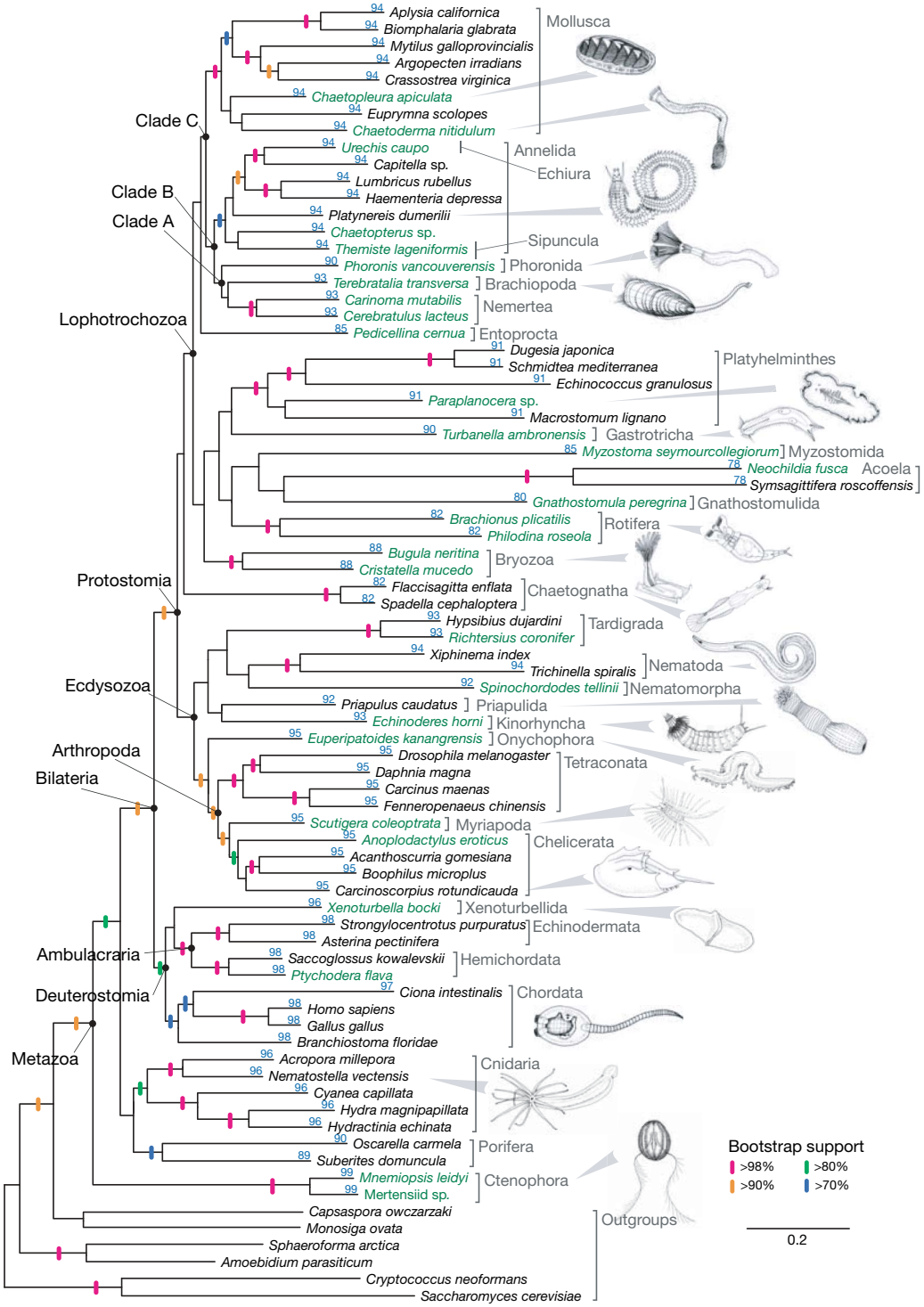


Figure 1 | Phylogram of the 77-taxon RaxML maximum likelihood analyses conducted under the WAG model. The figured topology and branch lengths are for the sampled tree with the highest likelihood (1,000 searches, log likelihood = –796,399.2). Support values are derived from 1,000 bootstrap replicates. Leaf stabilities are shown in blue above each branch. Taxa for which we collected new data are shown in green.

be that Platyzoa is an artefact of attracting unstable long-branch species to their vicinity.

Analyses of the 64-taxon matrix (Fig. 2 and Supplementary Fig. 9) show strong support for several important clades. To test if confidence in the relationships between stable taxa is overestimated in the absence of unstable taxa, we pruned away the 13 unstable taxa from each of the 1,000 bootstrap trees inferred from the 77-taxon matrix. This generated a set of trees containing only stable taxa, but for which relationships had been inferred in the presence of unstable taxa. Clade frequencies were calculated from this pruned tree set and mapped onto the most probable 64-taxon tree (Fig. 2). These reduced-tree support values are very similar to bootstrap support values calculated from the 64-taxon matrix, indicating that unstable taxa do not affect the inference of most relationships between stable taxa, only obscure these affinities.

The 64-taxon matrix strongly supports a sister-group relationship between Platyhelminthes and the remaining lophotrochozoans. A similar result, although uniting gastrotrichs with platyhelminths, was proposed recently[19]. Consistent with recent findings[20], *Urechis caupo*, an echiuran, is placed as sister to the annelid *Capitella* sp., and the sipunculan *Themiste lageniformis* is allied with annelids rather than molluscs. All analyses place Annelida as sister to a novel group that we call Clade A (Fig. 2), consisting of the nemerteans, a phoronid and a brachiopod, with variable support across analyses. Bayesian support for a group consisting of Annelida + Clade A (Clade B, Fig. 2) is strong (100% posterior probability in CAT and WAG analyses), whereas bootstrap support is moderate (84%). Although a brachiopod–annelid relationship is supported by the shared presence of chitinous chaetae, this new relationship implies that chaetae have been lost in nemerteans and phoronids (as in sipunculans, leeches

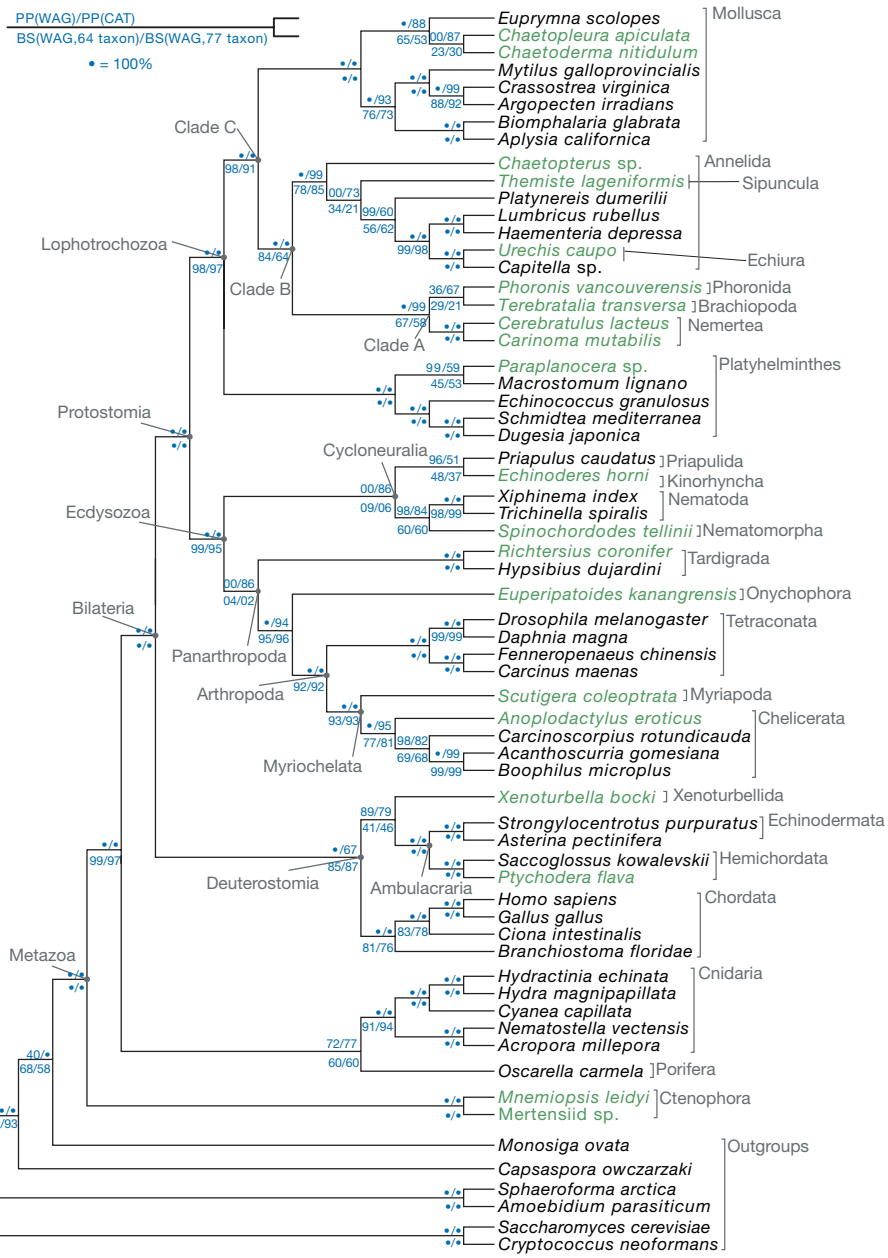


Figure 2 | Cladogram of the 64-taxon PhyloBayes bayesian analyses conducted under the CAT model. Posterior probabilities (PP) estimated under the CAT (15 PhyloBayes runs of 6,000 generations each; 1,200 generation burn-in) and WAG+I+Γ (8 MrBayes runs of two-million generations each; 125,000 generation burn-in; 4 chains per run) models. Maximum likelihood bootstrap support was calculated for the 64-taxon data set (2,000 replicate RaxML runs) and for the relationship of these 64 taxa in the 77-taxon analysis (by pruning all other taxa from the bootstrap replicates summarized in Fig. 1). Taxa for which we collected new data are shown in green. Support values, as specified at the top-left of the figure, are shown in blue.

and some other annelids). A monophyletic Mollusca, recovered here with significant support for the first time[21], is found to be sister to Clade B. Mollusca + Clade B (Clade C, Fig. 2) unites animals that produce chitinous chaetae with those that secrete $CaCO_3$ spicules and/or shells (that is, epidermal extracelluar formations for which secretory cells develop into a cup/follicle with microvilli at their base). A palaeontological scenario[22] identifies mollusc spicules and annelid/brachiopod chaetae as having been derived from distinctive fossil 'coelosclerites'. This scenario and a single origin of these epidermal formations are consistent with our cladogram.

The inclusion for the first time of nematomorphs, onychophorans and kinorhynchs in a phylogenomic analysis provides important insight into the structure of Ecdysozoa. Maximum likelihood bootstrap support for relationships within Ecdysozoa are similar in the 64- and 77-taxon analyses. The onychophoran is unambiguously placed as sister to arthropods in a clade of coelomate ecdysozoans that excludes Tardigrada, resolving a long-standing issue about the arthropods' sister group[5]. Tardigrades have traditionally been hypothesized to be allied with arthropods and onychophorans (together forming Panarthropoda)[23], but recent molecular data have suggested an alternative grouping of tardigrades with nematodes[9]. We find that the CAT model favours the former hypothesis (with Tardigrada sister to Onychophora + Arthropoda) whereas WAG favours the latter, indicating that at least one of these models is prone to systematic error for this particular problem (see Supplementary Information for further discussion of this issue).

We find strong support at all key internal arthropod nodes, and several contentious relationships of central interest are well resolved for the first time. Pycnogonids (sea spiders) group with chelicerates, rejecting placement of sea spiders as the earliest branching arthropod lineage[24]. Our results reject Mandibulata (Myriapoda, Crustacea and Hexapoda) in favour of myriapods being sister to chelicerates plus pycnogonids[25,26].

The spiral cleavage programme, a complex and highly stereotyped mode of early embryonic development, is present in at least Annelida, Entoprocta, Mollusca, Nemertea and Platyhelminthes[23], constituting a synapomorphy of at least the lophotrochozoan taxa included in the 64-taxon analysis. The placement of the lophophorate taxa Phoronida and Brachiopoda, which have radial cleavage and lie well within this assemblage, implies that they have lost spiral cleavage and also that their larvae are derived from the trochophore found in annelids, nemerteans and molluscs. Although phoronids do not show spiral cleavage, their mesoderm has a dual ecto/endodermal origin[27]—an important characteristic of spiralian embryology. Spiral cleavage has also been lost in cephalopod molluscs and in some neoophoran platyhelminths[23], establishing that this major shift has occurred repeatedly. Spiral cleavage may also have been lost or extensively modified in some of the unstable taxa not considered in the 64-taxon analysis (for example, gastrotrichs).

The placement of ctenophores (comb jellies) as the sister group to all other sampled metazoans is strongly supported in all our analyses. This result, which has not been postulated before, should be viewed as provisional until more data are considered from placozoans and additional sponges. If corroborated by further analyses, it would have major implications for early animal evolution, indicating either that sponges have been greatly simplified or that the complex morphology of ctenophores has arisen independently from that of other metazoans. Independent analyses of ribosomal and non-ribosomal proteins (Supplementary Information and Supplementary Fig. 10) indicate that support for this hypothesis (and for others presented for the first time here, such as Clade A and Clade B) is much greater in the combined analyses than in partitioned analyses with fewer genes. This may explain why these novel clades have not been recovered before, because support requires very broad gene sampling.

A few other principal groups have yet to be incorporated into phylogenomic studies, including Nemertodermatida, Loricifera, Cycliophora and Micrognathozoa. On the basis of our present findings, we predict that resolution across the metazoan tree will continue to improve as phylogenomic data from these additional taxa are collected and sampling is improved within clades already represented.

METHODS SUMMARY

Complementary DNA libraries were prepared for 29 species, and about 3,000 clones 5′ sequenced from each (Supplementary Table 1). All of our original sequence data have been deposited in the NCBI Trace Archive. These ESTs were assembled into a set of unique transcripts for each species, which were then translated into proteins using similarity and extension. Data from 48 additional species were downloaded from public archives (Supplementary Table 2). We present a new approach to identification of orthologous genes in animal phylogenomic studies (Supplementary Fig. 2) that relies on a Markov cluster algorithm[28,29] to analyse the structure of BLAST hits to a subset of the NCBI HomoloGene Database. The stringency of clustering is adjusted by means of the inflation parameter to best recapitulate the orthology groupings of HomoloGene.

Phylogenetic trees were inferred with bayesian and maximum likelihood approaches. The stabilities of taxa were assessed with leaf stabilities[17], as calculated by Phyutility[30] (available at http://code.google.com/p/phyutility/). Unstable taxa were removed from both sequence matrices and tree sets to assess the relationships of a stable subset of taxa to each other.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 September; accepted 20 December 2007.
Published online 5 March 2008.

1. Giribet, G. Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion? *Mol. Phylogenet. Evol.* **24,** 345–357 (2002).
2. Halanych, K. M. The new view of animal phylogeny. *Ann. Rev. Ecol. Evol. Sys.* **35,** 229–256 (2004).
3. Aguinaldo, A. M. A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387,** 489–493 (1997).
4. Halanych, K. M. *et al.* Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* **267,** 1641–1643 (1995).
5. Schmidt-Rhaesa, A. Tardigrades - Are they really miniaturized dwarfs? *Zool. Anz.* **240,** 549–555 (2001).
6. Philippe, H. & Telford, M. J. Large-scale sequencing and the new animal phylogeny. *Trends Ecol. Evol.* **21,** 614–620 (2006).
7. Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444,** 85–88 (2006).
8. Delsuc, F. *et al.* Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439,** 965–968 (2006).
9. Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* **22,** 1246–1253 (2005).
10. Philippe, H. *et al.* Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* **21,** 1740–1752 (2004).
11. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21,** 1095–1109 (2004).
12. Philip, G. K., Creevey, C. J. & McInerney, J. O. The Opisthokonta and the Ecdysozoa may not be clades: stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. *Mol. Biol. Evol.* **22,** 1175–1184 (2005).
13. Rokas, A., Kruger, D. & Carroll, S. B. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310,** 1933–1938 (2005).
14. Baurain, D., Brinkmann, H. & Philippe, H. Lack of resolution in the animal phylogeny: closely spaced cladogenesis or undetected systematic errors? *Mol. Biol. Evol.* **24,** 6–9 (2006).
15. Philippe, H. *et al.* Acoel flatworms are not Platyhelminthes: evidence from phylogenomics. *PLoS One* **2,** e717 (2007).
16. Blair, J. E. *et al.* The evolutionary position of nematodes. *BMC Evol. Biol.* **2,** 1–7 (2002).
17. Thorley, J. L. & Wilkinson, M. Testing the phylogenetic stability of early tetrapods. *J. Theor. Biol.* **200,** 343–344 (1999).
18. Giribet, G., Distel, D. L., Polz, M., Sterrer, W. & Wheeler, W. C. Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. *Syst. Biol.* **49,** 539–562 (2000).
19. Telford, M. J., Wise, M. J. & Gowri-Shankar, V. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. *Mol. Biol. Evol.* **22,** 1129–1136 (2005).
20. Struck, T. H. *et al.* Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evol. Biol.* **7,** 57 (2007).

21. Giribet, G. *et al.* Evidence for a clade composed of molluscs with serially repeated structures: monoplacophorans are related to chitons. *Proc. Natl Acad. Sci. USA* **103,** 7723–7728 (2006).
22. Conway Morris, S. & Peel, J. S. Articulated Halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Phil. Trans. R. Soc. Lond. B* **347,** 305–358 (1995).
23. Nielsen, C. *Animal Evolution, Interrelationships of the Living Phyla* 2nd edn (Oxford Univ. Press, Oxford, 2001).
24. Giribet, G., Edgecombe, G. D. & Wheeler, W. C. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413,** 157–161 (2001).
25. Mallatt, J. M., Garey, J. R. & Shultz, J. W. Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Mol. Phylogenet. Evol.* **31,** 178–191 (2004).
26. Hwang, U. W. *et al.* Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* **413,** 154–157 (2001).
27. Freeman, G. & Martindale, M. Q. The origin of mesoderm in phoronids. *Dev. Biol.* **252,** 301–311 (2002).
28. van Dongen, S. A cluster algorithm for graphs. *National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam.* Technical Report INS-R0010 (Stichting Mathematisch Centrum, Amsterdam, 2000).
29. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).
30. Smith, S. A. & Dunn, C. W. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics* doi:10.1093/bioinformatics/btm619 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all participants in the Protostome Assembling the Tree of Life (AToL) Project as well as E. J. Edwards, T. Dubuc, A. Stamatakis, J. Q. Henry and S. Maslakova. A.H. received support from the Deutsche Forschungsgemeinschaft, and M.O. received support from the Swedish Taxonomy Initiative and the Royal Swedish Academy of Sciences. The *Capitella* sp. EST data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/Capitella), as were the *Mnemiopsis* dbEST (http://www.ncbi.nlm.nih.gov/dbEST/) data. This work was funded by two consecutive collaborative grants from the AToL program from the US National Science Foundation. Ctenophore sequencing was supported by NASA.

Author Information The concatenated sequence matrix has been deposited at TreeBase (http://www.treebase.org). The raw sequence data are available at the NCBI Trace Archives (http://www.ncbi.nlm.nih.gov/Traces), and can be retrieved with the query 'center_name='KML-UH''. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.W.D. (casey_dunn@brown.edu).

METHODS

Molecular techniques. Total RNA was prepared using TRIzol (Molecular Research Center), the RNeasy Mini Kit (Qiagen), the RNAqueous-micro kit (Ambion) or Dynabeads (Invitrogen) from fresh specimens or tissue that had been stored in RNA*later* (Ambion) at –20 °C. First-strand cDNA was synthesized using the GeneRacer Kit (Invitrogen), which selects for full-length mRNA. Twenty cycles of PCR with the GeneRacer 5′ and 3′ primers were then performed (94 °C for 30 s, 69 °C for 30 s, and 72 °C for 4 min, with an initial denaturation of 94 °C for 5 min and a final extension of 72 °C for 10 min; BD Advantage 2 Polymerase Mix, Clontech). The PCR products of most taxa were enriched for larger fragments using ChromaSpin TE400 columns (Clontech). PCR products were concentrated with the MinElute PCR Purification Kit (Qiagen) and ligated into pGEM-T Easy (Promega). The ligations were sent to Macrogen Ltd for transformation, plating, colony picking, minipreping, and 5′ sequencing with the GeneRacer 5′ primer. All of our original sequence data have been deposited in the NCBI Trace Archive.

Sequence preprocessing. The PartiGene Pipeline v3.0 (ref. 31) was used to preprocess EST data, with several modifications (Supplementary Fig. 2). The option to use quality data for assembly was enabled. Partigene outputs multiple contiguous sequences for a given transcript when PHRAP (http://www.phrap.org/) does not fully assemble the sequences assigned to a transcript. Low-quality ends were trimmed from these partially assembled sequences, which were then aligned with ClustalW[32] and the highest-quality bases chosen for the consensus. Transcripts were translated by similarity and extension (using the SwissProt database).

The 2,137 *Xenoturbella bocki* sequences from dbEST were assembled along with the 3,840 new sequences that we generated. The 3,360 ESTs we prepared from *Mnemiopsis leidyi* were also combined with data from dbEST that had been generated by the US Department of Energy Joint Genome Institute. In addition, we considered 48 taxa from other publicly available sources (Supplementary Table 2).

Orthology assignment. We developed an explicit method for selecting genes from EST data sets to maximise gene intersection across taxa and to minimise problems with orthology and paralogy (Supplementary Fig. 2). Promiscuous domains (Conserved Domain Database[33] accession numbers pfam01535, pfam00400, pfam00047, smart00407, cd00099, pfam00076, pfam00023, pfam01576, pfam00041, cd00031, smart00112, cd00096, cd00204, pfam00023, smart00248, pfam01344, pfam00018, pfam00038, pfam00096, pfam00595, pfam00651, pfam00169, pfam00105, pfam00435, pfam00084, pfam00017, smart00225, smart00367, smart00135, cd00020, pfam00514, cd00020, smart00185, cd00014, pfam00307 and smart00033) were identified by RPSBLAST and masked before orthology assignment. These domains are a subset of those masked in the construction of NCBI KOG database of eukaryotic orthologues[34]. We constructed a local database of all *Homo sapiens*, *Canis familiaris*, *Gallus gallus*, *Drosophila melanogaster* and *Anopheles gambiae* sequences that have orthology assignments in the National Center for Biotechnology Information (NCBI) HomoloGene database, and the masked sequences were queried against these sequences with BLASTP. BLASTP hits were then passed to TribeMCL (the version bundled with mcl v6.58) for Markov Chain Clustering (MCL)[29,35]. The MCL inflation parameter was varied in intervals of 0.1 to identify the value that generated the maximum number of clusters with sequences from one HomoloGene group.

Groups with sequences from fewer than 25 taxa were discarded. We also discarded groups with sequences from fewer than 5 of the taxa we collected original EST data for to prevent gene selection from being dominated by some of the much larger EST and genomic data sets included from public archives. The number of sequences for each taxon represented within each group was then enumerated, and groups with a median of greater than one or a mean greater than 2.5 were discarded. This eliminated many groups that had a high rate of lineage-specific duplication. Two features of the cluster graph were then evaluated for properties potentially indicative of paralogy problems. First, the group was rejected if it included no Homologene sequences. Second, the TribeMCL group was rejected if it included any Homologene sequences belonging to a Homologene group with sequences in another TribeMCL group.

Most TribeMCL groups contained multiple sequences for some taxa, which could be paralogues, splice variants or the result of EST assembly errors. The sequences for each of these problematic TribeMCL groups were aligned with ClustalW v1.83 (ref. 32), and parsimony trees (100 bootstrap replications) were inferred with PAUP* v4.0b10 (ref. 36). All but one of the sequences from the same taxon were automatically excluded from the group if they were monophyletic with a bootstrap score of >80%. The retained sequence was selected to have a stop codon if possible. Trees for TribeMCL groups that still had taxa with multiple sequences were then visually inspected. If there were strongly supported deep nodes indicating the existence of multiple paralogues shared by multiple taxa the entire group was excluded. Otherwise, all sequences for the problematic taxa were excluded from the group and sequences from nonproblematic taxa retained.

All groups that passed the above criteria were prepared for tree building. 5′ untranslated regions were removed by blasting each sequence against the other sequences in the same group and trimming ends that were not included in the resulting HSPs (10^{-4} *e*-value threshold). The sequences of each TribeMCL group were aligned with Muscle v3.6 (ref. 37) and trimmed with Gblocks v0.91b[38] (settings: –b2 = [65% of the number of sequences] –b3 = 10 –b4 = 5 –b5 = a). These trimmed alignments for each gene were then concatenated into a single alignment (21,152 positions long), which has been deposited in TreeBase.

To compare matrix construction methods between studies, sequences were queried by BLASTP (10^{-20} *e*-value threshold) against the sequences of the most frequently used matrix of genes in metazoan EST studies[9]. The identity of the top-scoring hit, if any hits were found, was putatively assigned to the query sequence. Alignment and trimming were executed as described above, and the least-divergent sequences were assembled into a matrix (24,708 positions long) with SCaFoS[39].

Phylogenetic analyses. Phylogenetic analysis of our large matrix was computationally intensive and took several months on more than 120 processors spread across multiple modern computer clusters. A preliminary matrix was evaluated under a mixed model with MrBayes v.3.1.2 (ref. 40), which selected WAG with 100% posterior probability. Maximum likelihood analyses were performed with RAxML-VI-HPC v.2.2.1 (ref. 41). All searches were completed with the PROTMIXWAG option. PhyloBayes v.2.1 (ref. 11) was used for bayesian analyses conducted under the CAT model, and MrBayes v.3.1.2 for bayesian analyses under the WAG model (with Gamma approximation of among site rate variation and allowing for invariable sites). Burn-ins were determined by plotting parameters across all runs for a given analysis. Leaf stabilities[17] were calculated with the tree analysis program Phyutility[30] (available at http://code.google.com/p/phyutility/), which was also used to determine where unstable taxa wandered across the bootstrap replicates (Supplementary Fig. 8).

31. Parkinson, J. *et al.* PartiGene — constructing partial genomes. *Bioinformatics* **20,** 1398–1404 (2004).
32. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
33. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33** (Database issue), D192–D196 (2005).
34. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4,** 41 (2003).
35. van Dongen, S. *Graph Clustering by Flow Simulation*. PhD thesis, Univ. Utrecht (2000).
36. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods)* Version 4 (Sinauer Associates, Sunderland, Massachusetts, 2003).
37. Edgar, R. C. & Journals, O. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
38. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17,** 540–552 (2000).
39. Roure, B., Rodriguez-Ezpeleta, N. & Philippe, H. SCaFoS: a tool for Selection, Concatenation and Fusion of Sequences for phylogenomics. *BMC Evol. Biol.* **7,** S2 (2007).
40. Huelsenbeck, J. P. & Ronquist, F. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17,** 754–755 (2001).
41. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).

LETTERS

Estimating the impact of school closure on influenza transmission from Sentinel data

Simon Cauchemez[1], Alain-Jacques Valleron[2,3,4], Pierre-Yves Boëlle[2,3,4], Antoine Flahault[2,3,5] & Neil M. Ferguson[1]

The threat posed by the highly pathogenic H5N1 influenza virus requires public health authorities to prepare for a human pandemic. Although pre-pandemic vaccines and antiviral drugs might significantly reduce illness rates[1,2], their stockpiling is too expensive to be practical for many countries. Consequently, alternative control strategies, based on non-pharmaceutical interventions, are a potentially attractive policy option. School closure is the measure most often considered. The high social and economic costs of closing schools for months make it an expensive and therefore controversial policy, and the current absence of quantitative data on the role of schools during influenza epidemics means there is little consensus on the probable effectiveness of school closure in reducing the impact of a pandemic. Here, from the joint analysis of surveillance data and holiday timing in France, we quantify the role of schools in influenza epidemics and predict the effect of school closure during a pandemic. We show that holidays lead to a 20–29% reduction in the rate at which influenza is transmitted to children, but that they have no detectable effect on the contact patterns of adults. Holidays prevent 16–18% of seasonal influenza cases (18–21% in children). By extrapolation, we find that prolonged school closure during a pandemic might reduce the cumulative number of cases by 13–17% (18–23% in children) and peak attack rates by up to 39–45% (47–52% in children). The impact of school closure would be reduced if it proved difficult to maintain low contact rates among children for a prolonged period.

A thorough evaluation of the effectiveness of school closure as a pandemic mitigation measure is difficult, owing to the limited epidemiological data[3] and the current deficit in statistical methods to analyse those data. So far, it has been possible to establish that school closure is negatively correlated with influenza incidence[4,5]. Mathematical models have been used to evaluate the impact of school closure in a pandemic[1,2,6]. However, in the absence of quantitative estimates derived from epidemiological data, those models made strong assumptions about school transmission. The relatively wide range of effects they predicted[1,2,6] shows that modelling assumptions cannot replace the statistical investigation of epidemiological data.

Here we present a novel statistical approach to evaluating the impact of school closure on influenza epidemics from the joint analysis of disease surveillance data and information on the timing of school holidays in France. The hypothesis we examine is that influenza transmission changes during holidays as a result of the altered mixing patterns of children. The Sentinel network[7,8] (see http://www.sentiweb.org and Supplementary Information) is an internet-based network of French general practitioners (GPs). Since 1984, approximately 1,200 GPs have collected and sent data regularly on a dozen diseases, including influenza-like illness (case definition: sudden temperature of $>39\,^{\circ}C$, myalgia and cough/running nose). Regional daily incidences of influenza-like illness are estimated as area-weighted averages from individual GP declarations, using population data and data on the percentage of GPs participating in the surveillance network. Data on the timing of French holidays in different regions was obtained from the French Ministry of Education. In France, holidays are staggered across three geographic zones (two zones in 1986 and 1990) and the timing varies from region to region and from year to year. This provides conditions resembling those of a natural experiment.

The surveillance data consist of daily incidences for children (<18 years old) and adults (≥ 18 years old) for the two or three holiday zones in mainland France and over 21 years (1985–2006). We assume that half of all influenza patients consult his or her GP, giving a reasonable average attack rate of 11.4% (range 4.6–20.6%). We select epidemic periods (weekly incidence >160 per 100,000 inhabitants) and discard one epidemic that lasted 13 days only. This leaves 60 epidemic periods, with average duration 61 days (range 22–111 days) (Fig. 1a).

We model the spread of influenza in a population structured into households and schools (Fig. 1d; see Methods and Supplementary Information). Community transmission also occurs randomly between all members of the population. The simulated population matches the structure of the French population (Fig. 1b, c). We assume that at the start of each influenza season an average of 27% of the population is immune[9], and that immunity is distributed within the population from its expected stationary distribution (see Supplementary Information). During holidays, no transmission occurs in schools, but in other places (household, community), transmission rates may be modified. We use estimates from another study[10] to characterize household transmission and the infectiousness profile (generation time 2.4 days).

The high dimensionality of the data means that model parameters cannot be estimated using standard statistical methods, such as least-squares fitting or data augmentation[10]. We therefore designed a new statistical approach, based on the simulation of epidemics that are constrained to be consistent with the observed incidence curves (Fig. 1e; see Methods and Supplementary Information). Using simulated data, we find that, even in a context with observation errors and where transmissibility varies substantially between epidemics, the inference method gives satisfactory estimates of all parameters (see Supplementary Information). The approach also provides the relative prediction error (RPE; see Fig. 2a).

We first estimate transmission parameters under the assumption that influenza transmission is not modified during holidays (see Supplementary Information). For this model, adult and child RPEs are close to 0% during the school term (Fig. 2a). Adult RPE is also close to 0% during holidays; but child RPE drops to -24% (range -20% to -29%) during holidays. This implies that, on average,

[1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Diseases Epidemiology, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Université Pierre et Marie Curie—Paris 6, UMR S 707, 27 rue Chaligny, Paris 75012, France. [3]INSERM, UMR S 707, 27 rue Chaligny, Paris 75012, France. [4]AP-HP, Hôpital St Antoine, 27 rue Chaligny, Paris 75012, France. [5]French School of Public Health (EHESP), 1 place du Parvis Notre-Dame, Paris F-75004, France.

holidays lead to a 24% reduction in the rate at which influenza is transmitted to children, but that they have no detectable effect on adults' contact patterns.

We then estimate transmission parameters allowing for modifications of children's contact patterns during holidays. Three quantities are needed: transmission rates within schools, in the community, and the increase in non-school transmission rates during holidays (compensatory behaviours). It is not possible to estimate those three quantities independently from the available data. We therefore fix one of them (compensatory behaviours) and estimate the other two. We then undertake a rigorous sensitivity analysis on compensatory behaviours, considering 30 parameter combinations parameterized by the increase in child-to-child community transmission ($\delta_{com} = 0\%$, 50%, 100%, 150%, 200% and up to ∞) and in child-to-child household transmission ($\delta_{hous} = 0\%$, 50%, 100%, 150% and 200%).

Irrespective of $\{\delta_{com}, \delta_{hous}\}$, we find that accounting for holidays improves the model fit: (1) log-likelihoods are larger (see Supplementary Information); (2) child RPE becomes close to 0% during holidays (Fig. 2a). We find that the proportion of transmission occurring in schools increases with $\{\delta_{com}, \delta_{hous}\}$, ranging from 7 to 20% overall (Fig. 2c) and from 20 to 54% in children (Fig. 2d). The proportion of secondary cases of children infected in schools is 16–44%. Other summary statistics are robust to a change in $\{\delta_{com}, \delta_{hous}\}$ (see Supplementary Information). Although children represent 28% of the population, they are responsible for 46–47% of all infections (Fig. 2e). Household transmission accounts for 36–39% of infections of children (and 40% of adult infections). Household members make up 48–50% of secondary cases in children. The basic reproduction number, R_0 (average number of cases generated by one typical case in a completely susceptible population) is estimated to be 1.7 (range 1.5–1.8) during school term, and 1.4 (range 1.3–1.6) in holidays. The average R_0 for child cases is 2.2 (range 2.0-2.4) during term and 1.7 (1.4–1.9) during holidays, while for adult cases it is 1.3 (range 1.2-1.4) for both terms and holidays (see Supplementary Information).

No major difference in transmission is detected between Christmas and other breaks (winter and spring breaks). For adults, RPE is 0% (range −4% to 5%) over Christmas and 1% (range −4% to 5%) during other breaks. For children, RPE is −4% (range −15% to 9%) over Christmas and −1% (range −12% to 12%) during other breaks.

Classifying each year by the dominant influenza virus type or subtype and fitting season-specific variations in transmissibility, we find that subtype B is less transmissible but more child-associated

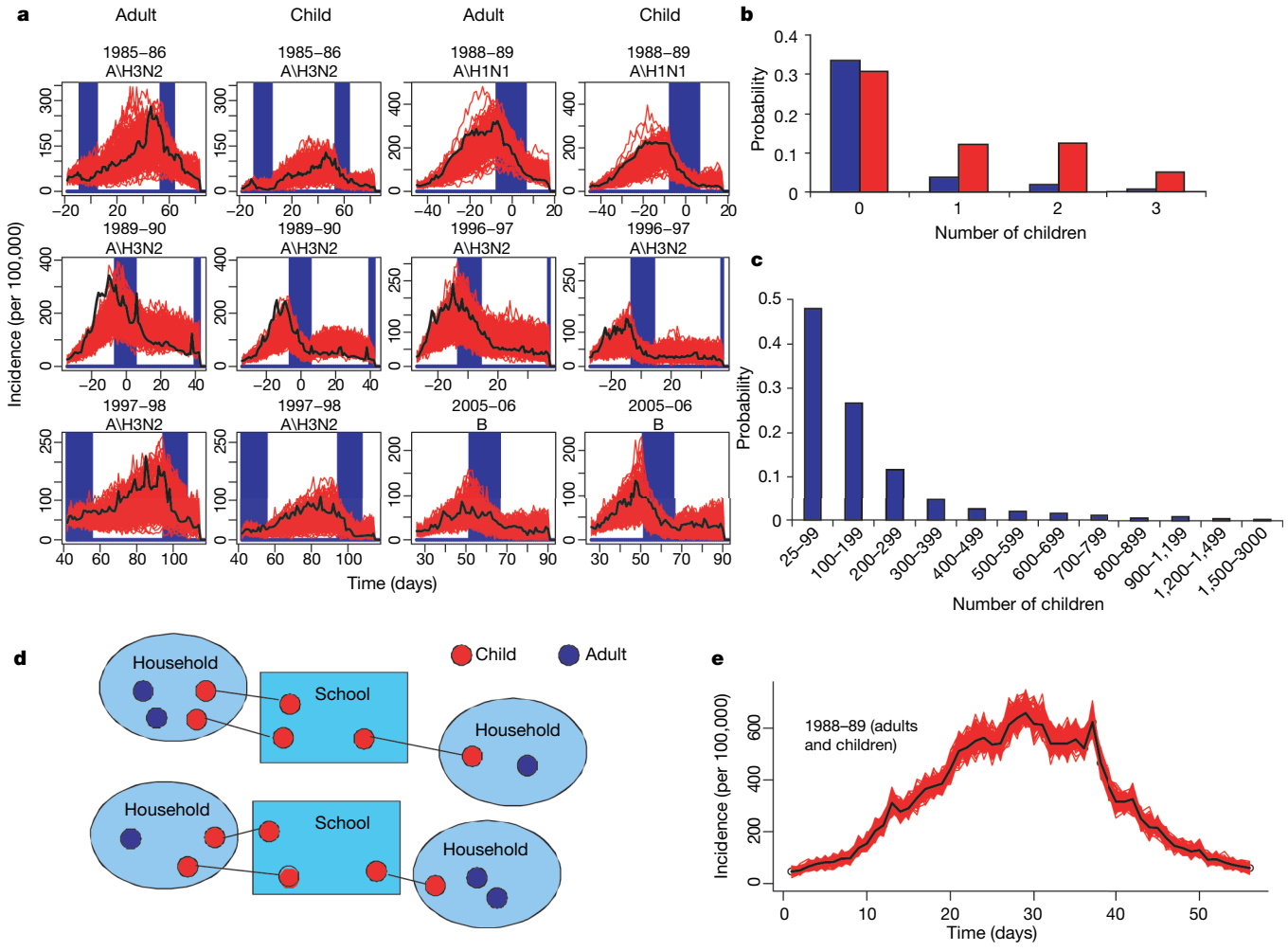


Figure 1 | Data, transmission model and inference method. **a**, Daily incidence (black line) for children (<18 years old) and adults and holiday timing (blue bars) for six of the 60 epidemic periods selected among surveillance data over 21 years (1985–2006) and three holiday zones in France. Day 0 corresponds to 1 January. Red lines show 200 simulations from the model, with parameters drawn from their posterior distribution. **b**. Size distribution of French households (1999 census). Blue, one adult; red, two adults. **c**, Size distribution of French schools (1999 census). **d**, Schematic diagram of transmission model in a population structured into households and schools (see Methods and Supplementary Information). **e**, Constrained simulations. For inference, epidemics are simulated which are constrained to be consistent with observed incidence curves. Black line, the observed incidence curve for one holiday zone in 1988–89; red lines, 200 constrained simulations. See Methods and Supplementary Information.

than subtype A\H3N2 (Wilcoxon test: probability $P = 2.6\%$ for the strength of transmission, and $P = 1.8\%$ for the relative contribution of children to transmission) and that subtype A\H1N1 has intermediate characteristics between subtype B and subtype A\H3N2 (no significant difference with subtype B, nor with A\H3N2) (Fig. 2b).

We then simulate epidemics from the model, with parameters drawn from their posterior distribution for the different model variants (Fig. 1a and Supplementary Information). Simulations start at the same time and with the same number of initial cases as observed epidemics, use season-specific transmissibility estimates, but are not otherwise constrained. For adults, 81% (83% for children) of observed daily attack rates fall between the 2.5th and 97.5th percentile of the distribution of simulated daily attack rates.

These simulations are used to assess the impact of school closure on seasonal and pandemic attack rates. Figure 3a–c shows that, irrespective of the assumptions made about $\{\delta_{com}, \delta_{hous}\}$, we obtain the same estimates of the impact of school closure on cumulative and peak attack rates. For typical holiday timings, the different model variants predict an average seasonal attack rate of 10.6–11.1%. They also predict that, if schools were always open, the attack rate would be 12.8–13.4%; that is, holidays prevent 16–18% of seasonal influenza cases (for adults 14–17%; for children 18–21%).

We then consider the pandemic context, where 100% of the population is susceptible and assume that 50% of infections are symptomatic. For typical holiday timings, 31% of the population would report being ill (37–38% of children); and the daily incidence at the peak would be 1.6–1.7% (2.1–2.2% in children). If schools were closed permanently at an early stage (for example, once daily incidence exceeds 20/100,000), with subsequent behaviour typical of holidays, the cumulative number of cases would be curbed by 13–17% overall (for children only 18–23%) and the number of cases at the peak by 39–45% (for children only 47–52%).

Contact patterns might, however, be less affected by prolonged school closure than by normal school breaks, when people go on vacation, celebrate Christmas, and so on. The reductions we predict might therefore be an upper bound of what might happen during school closure in a pandemic. If compensatory increases in contact rates $\{\delta_{com}, \delta_{hous}\}$ were 1.5-fold larger during school closure in a pandemic than for typical holidays, there would be at most a very limited reduction in cumulative/peak attack rates and in R_0 (Fig. 3d–i and Supplementary Information).

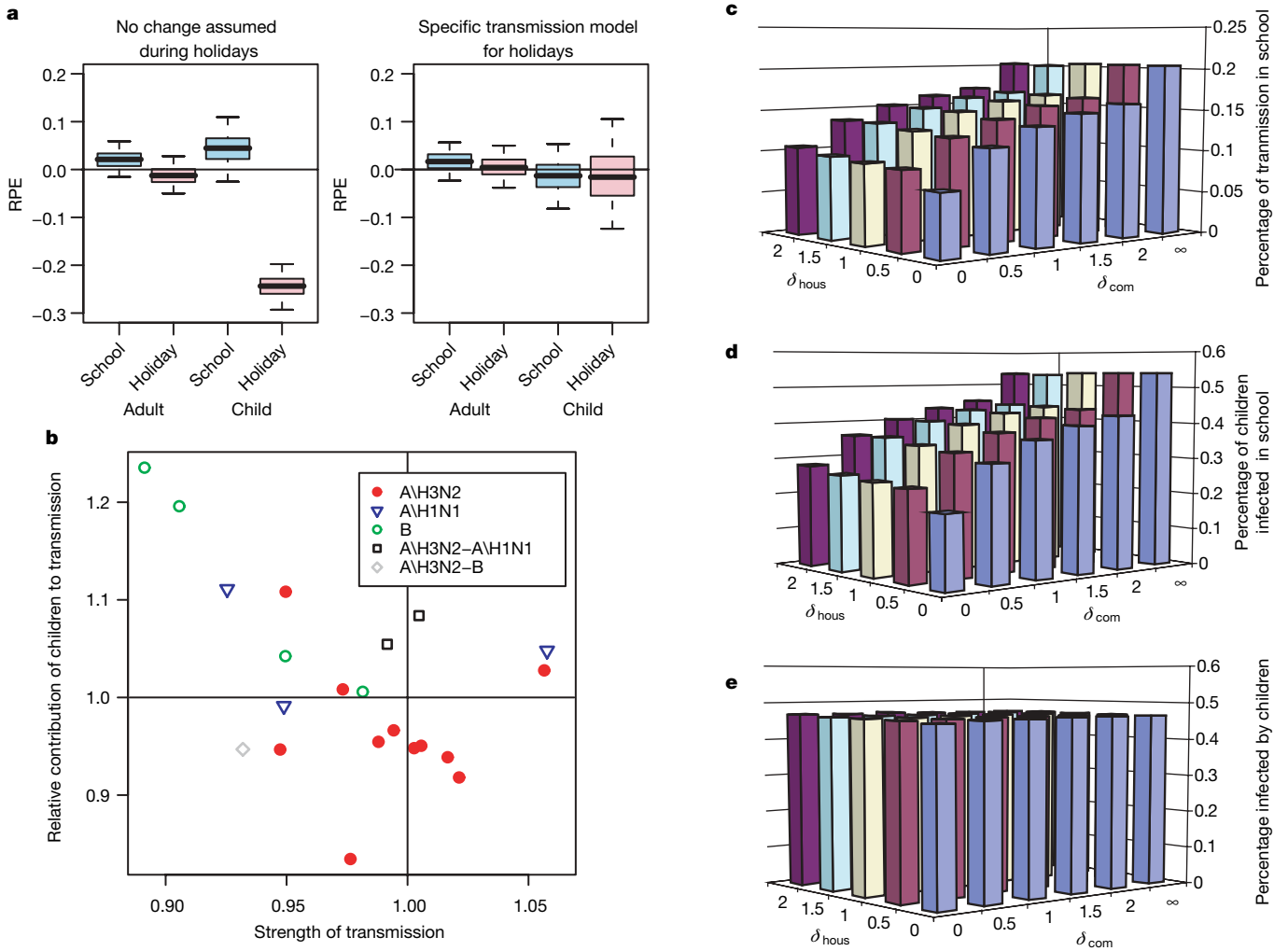


Figure 2 | Inferred influenza transmission characteristics. **a**, Posterior distribution (2.5%, 25%, 50%, 75% and 97.5% percentiles) of the RPE for adults and children, during school terms and during holidays, when no change in transmission is assumed during holidays (left) and when a specific transmission model is designed for holidays (right). RPE is the average relative error $(O_t - E_t)/E_t$ between the number O_t of cases observed at time t and the number E_t of cases predicted by the model given the observed epidemic up to time $t - 1$. An RPE close to 0% is indicative of good fit. **b**, Strength of transmission, and relative contribution of children to transmission for each epidemic season according to the circulating subtype[19] (see Methods and Supplementary Information for details of calculation). **c**, Proportion of school transmission according to the strength of compensatory behaviours in the community (δ_{com}) and in the household (δ_{hous}). **d**, Proportion of children infected in schools. **e**, Proportion of cases infected by children.

For one model variant ($\delta_{com} = \delta_{hous} = 100\%$), we then perform a sensitivity analysis and find that results are relatively robust to varying other model assumptions (Fig. 3d–i and Supplementary Information). If prolonged school closure has the same impact as holidays, the relative reduction in the cumulative number of cases in a pandemic is always below 20%, except under the unlikely (see Supplementary Information) assumptions that immunity is completely clustered in households (42% reduction) or that the generation time is as long as 4.1 days (25% reduction). Summary statistics on the place of transmission are also relatively robust to changes in modelling assumptions (see Supplementary Information).

The derivation of influenza incidence from GP reports is uncertain because some cases do not visit a GP, only a small proportion of GPs report, diagnosis is based on influenza-like illness with no viral ascertainment and there are asymptomatic infections. Three observations do however suggest that the influenza-like illness data provide a sensible description of influenza circulation. First, they are consistent with data collected independently on virus circulation (see Supplementary Information). Second, weekly mortality due to pneumonia and influenza is almost perfectly predicted by the surveillance data and the circulating strains[11]. Lastly, solely on the basis of the influenza-like illness data, we found that transmission characteristics of influenza depended on the circulating subtype (Fig. 2b), in a way that is consistent with past epidemiological studies[12–14].

Although the apparent impact of public health measures was substantial (that is, up to 50% reduction in transmission) in some US cities in 1918, it is not possible to disentangle the relative impact of different measures[15–17]. School closure was commonly adopted, and in some of the cities in which schools were closed, the total impact of all public health measures was estimated to be as low as 10% (ref. 15). Here, we used a natural experiment to estimate the specific effect of school closure on seasonal influenza transmission. Our extrapolations to the pandemic context rest on the relatively strong assumption that people will behave during a pandemic as they do during seasonal outbreaks.

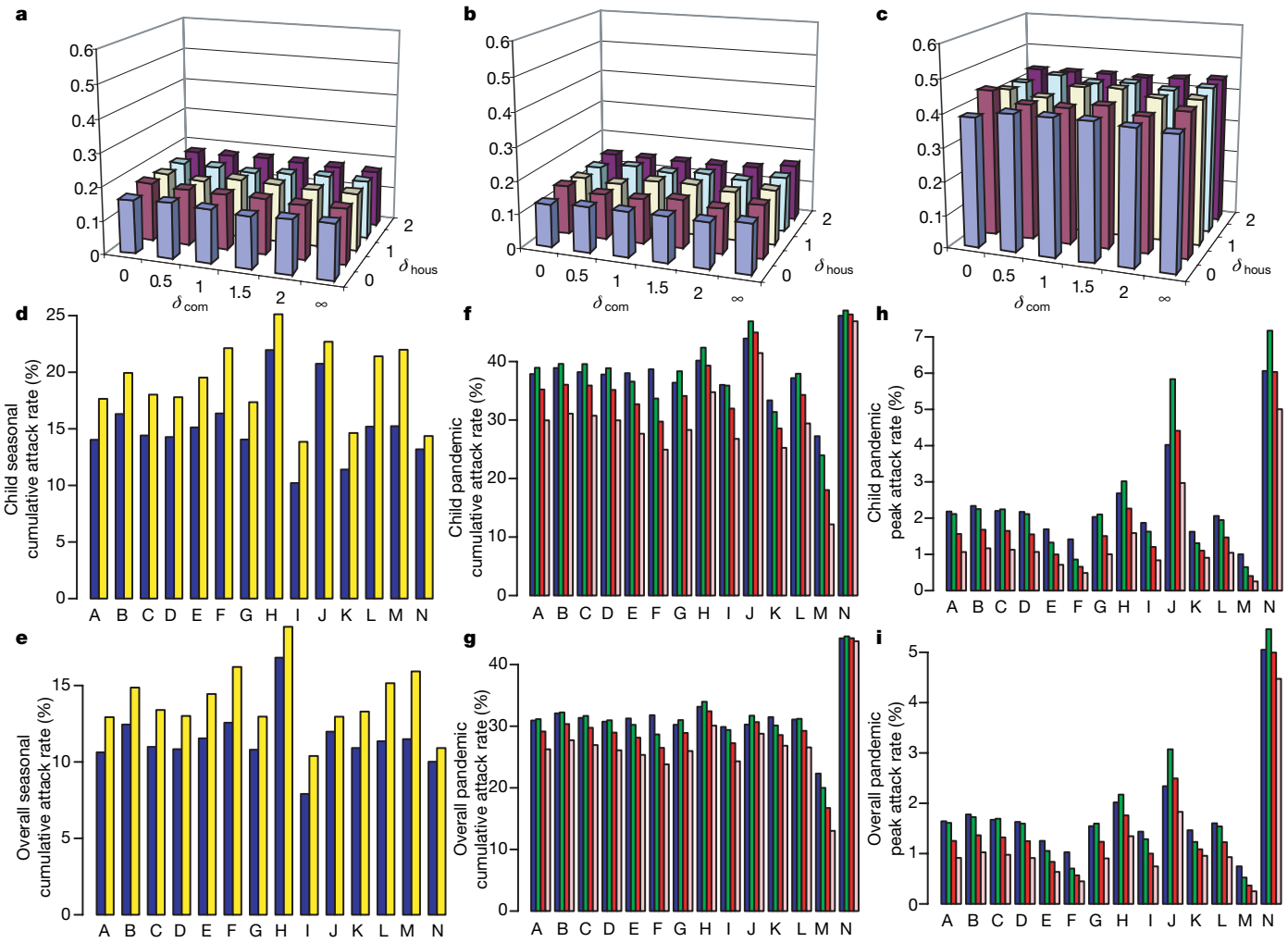


Figure 3 | Impact of school closure on seasonal and pandemic influenza. **a**, Relative reduction in seasonal influenza cumulative attack rates due to holidays, according to the assumed compensatory increase in contact rates in the community (δ_{com}) and in the household (δ_{hous}) during French holidays, and under baseline assumptions (see main text). **b**, Relative reduction in pandemic cumulative attack rates due to permanent school closure, assuming closure has the same effect on transmission as holidays. **c**, As for **b**, but for peak daily attack rate. **d–i**, Sensitivity analyses for parameters estimated assuming $\delta_{com} = \delta_{hous} = 100\%$ during French holidays. A: baseline (see main text); B: 19 smallest outbreaks discarded; C and D: epidemic period defined as weekly incidence over 120/100,000 or 200/100,000 respectively; E and F: 3.25-day and 4.11-day generation time respectively; G: household transmission rates 25% smaller than estimated in ref. 10; H–K: adult reporting rates of 30%, 70%, 50% and 50% respectively, with child reporting rates of 30%, 70%, 30% and 70% respectively; L: immunity seeded independently of household; M: immunity clustered by household; N: 50% immune. **d**, Seasonal cumulative attack rates among children during a typical holiday pattern (blue) and when schools are never closed (yellow). **e**, As for **d**, but for the whole population. **f**, Pandemic cumulative attack rates among children during a typical holiday pattern (blue); when schools are closed throughout with $\delta_{com} = \delta_{hous} = 100\%$ (pink); when schools are closed throughout but with compensatory contact rate increases 1.5-fold larger than normal holidays, that is $\delta_{com} = \delta_{hous} = 150\%$ (green); and as for green but with $\delta_{com} = \delta_{hous} = 125\%$ (red). **g**, As for **f**, but for the whole population. **h**, As for **f**, but for peak daily attack rates. **i**, As for **h**, but for the whole population.

Because demography and school holiday patterns are similar across much of Europe, we are confident that our results can be extrapolated to other European countries. Extrapolation to developing countries is more difficult because of the absence of independent data.

Compared with other studies of influenza transmission[18,19], our analysis shows that age is an important determinant of seasonal variations in influenza transmission both within (holidays fundamentally affect children's contact patterns) and between epidemics (there are large variations between seasons in the relative contribution of children to transmission).

Methods used to estimate parameters of complex transmission models (for example, data augmentation techniques[10]) have traditionally been very distinct from techniques used for prediction (that is, simulation[1]). The difficulty (and sometimes, as here, impossibility) of implementing those estimation methods for high-dimensional dynamical models largely explains why estimation often rests on naive least-squares fitting. In contrast, the new statistical method presented here, which relies on sequential Monte Carlo methods[20,21], makes it straightforward to upgrade a complex epidemic simulator to a computationally efficient likelihood-based inference tool.

Pandemic planning is a challenging task in today's highly connected world and when some key characteristics of the future pandemic virus cannot be predicted. Mathematical models provide a framework for assisting rational decision-making. However, for models to have predictive power, it is critical that they make full use of epidemiological data. Undertaking more epidemiological studies and designing statistical methods to extract maximum information from the data collected therefore remains a priority.

In public health terms, our conclusions do not rule out the use of school closure in a severe pandemic. We predict that this policy can significantly reduce the stress on healthcare systems at the peak of the pandemic. But our work should temper expectations of the scale of the reduction in overall illness and mortality achievable through this measure alone.

METHODS SUMMARY

Transmission model. The household transmission rate associated with an infective person of age a (where $a = $ A for adult or C for child) is $\beta^a_{\text{hous}} f(t)/n$, where n is the size of the household and $f(t)$ characterizes the relative infectiousness at time t since infection. An infectious child infects children in the same school at a rate $\beta_{\text{school}} f(t)/N_{\text{school}}$ where N_{school} is the size of the school. We make a distinction between adult-to-adult (A→A), child-to-child (C→C) and adult-to-child or child-to-adult (A→C, C→A) transmission in the community. During holidays, C→C community transmission increases by a factor of $1+\delta_{\text{com}}$ ($1+\delta_{\text{hous}}$ for household transmission). We explore a range of possible compensatory behaviours, parameterized by $\delta_{\text{com}} = 0\%$, 50%, 100%, 150% 200% and ∞ ($\delta_{\text{com}} = \infty$ is the extreme situation where children mix only in schools during school terms and mix only in the community during holidays) and $\delta_{\text{hous}} = 0\%$, 50%, 100%, 150% and 200%. At any time t, susceptible individual i is exposed to a baseline infection risk of $\lambda_i^{\text{baseline}}(t)$, which is the sum of the baseline risks of infection in their household, their school (for children) and the community. To model seasonal variations in influenza transmission, we introduce for each year y the strength of transmission σ_y and the relative contribution of children to transmission τ_y. During year y, at time t, the risk of infection $\lambda_{i,y}(t)$ for susceptible i is $\sigma_y \tau_y \lambda_i^{\text{baseline}}(t)$ if i is a child and $(\sigma_y/\tau_y)\lambda_i^{\text{baseline}}(t)$ if i is an adult. See the Supplementary Information for more information.

'Constrained' simulations. To approximate the likelihood of the parameters, we simulate epidemics constrained to be consistent with the observed incidence curves (Fig. 1e). At any time t, on average, the number of cases generated by the constrained simulator equals the observed number of cases. The likelihood is then approximated by sequential importance sampling[20,21], and the parameters space is explored by Markov-chain Monte-Carlo sampling[22,23]. Details on the simulation of constrained epidemics and the statistical methodology are given in the Supplementary Information and online-only Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2007; accepted 21 January 2008.

1. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442,** 448–452 (2006).
2. Germann, T. C., Kadau, K., Longini, I. M. Jr & Macken, C. A. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* **103,** 5935–5940 (2006).
3. Bell, D. M. Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerg. Infect. Dis.* **12,** 88–94 (2006).
4. Heymann, A., Chodick, G., Reichman, B., Kokia, E. & Laufer, J. Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. *Pediatr. Infect. Dis. J.* **23,** 675–677 (2004).
5. Valleron, A. J., Flahault, A. & Boelle, P. Y. *Do school holidays have an impact on influenza epidemics, then on mortality?* Presentation at *International Conference on Options for the Control of Influenza V* (Okinawa, 7–11 October 2003); ⟨http://www.u707.jussieu.fr/valleron/dia/dia2003/okinawa.pdf⟩.
6. Glass, R., Glass, L., Beyeler, W. & Min, H. Targeted social distancing design for pandemic influenza. *Emerg. Infect. Dis.* **12,** 1671–1681 (2006).
7. Valleron, A. J. *et al.* A computer network for the surveillance of communicable diseases: the French experiment. *Am. J. Public Health* **76,** 1289–1292 (1986).
8. Flahault, A. *et al.* Virtual surveillance of communicable diseases: a 20-year experience in France. *Stat. Methods Med. Res.* **15,** 413–421 (2006).
9. Longini, I. M. Jr, Koopman, J. S., Haber, M. & Cotsonis, G. A. Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. *Am. J. Epidemiol.* **128,** 845–859 (1988).
10. Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. & Boelle, P. Y. A. Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23,** 3469–3487 (2004).
11. Denoeud, L. *et al.* Predicting pneumonia and influenza mortality from morbidity data. *PLoS One* **2,** e464 (2007).
12. Glezen, W. P. *et al.* Age distribution of patients with medically-attended illnesses caused by sequential variants of influenza A/H1N1: comparison to age-specific infection rates, 1978–1989. *Am. J. Epidemiol.* **133,** 296–304 (1991).
13. Monto, A. S. & Sullivan, K. M. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol. Infect.* **110,** 145–160 (1993).
14. Olson, D. R. *et al.* Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. *PLoS Med.* **4,** e247 (2007).
15. Bootsma, M. C. J. & Ferguson, N. M. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc. Natl Acad. Sci. USA* **104,** 7588–7593 (2007).
16. Hatchett, R. J., Mecher, C. E. & Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl Acad. Sci. USA* **104,** 7582–7587 (2007).
17. Markel, H. *et al.* Nonpharmaceutical interventions implemented by US cities during the 1918–1919 influenza pandemic. *J. Am. Med. Assoc.* **298,** 644–654 (2007).
18. Finkenstadt, B. F., Morton, A. & Rand, D. A. Modelling antigenic drift in weekly flu incidence. *Stat. Med.* **24,** 3447–3461 (2005).
19. Xia, Y., Gog, J. R. & Grenfell, B. Semiparametric estimation of the duration of immunity from infectious disease time series: influenza as a case-study. *J. R. Stat. Soc. Ser. C* **54,** 659–672 (2005).
20. Doucet, A., de Freitas, N. & Gordon, N. *Sequential Monte Carlo Methods in Practice* (Springer, New York, 2001).
21. Liu, J. S. *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2001).
22. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the MRC, European Union FP6 SARSTRANS and INFTRANS projects, RCUK, and the NIGMS MIDAS initiative for research funding. We thank F. Carrat for comments.

Author Contributions S.C. developed the transmission model and conceived and implemented the inference framework used, did the analysis and drafted and revised the text. All other authors edited or commented on the text. A.-J.V., P.-Y.B. and A.F. identified, collated and processed the surveillance and holiday data. P.-Y.B. also provided input on the statistical framework. N.M.F. conceived the study (building on earlier work by A.-J.V. and A.F. examining the correlation between holidays and seasonal influenza incidence), provided input on the statistical framework, model design and assumptions and gave other technical advice.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.C. (s.cauchemez@imperial.ac.uk).

METHODS

Simulation of 'unconstrained' epidemics. Let θ be the baseline transmission parameters and V the set of parameters characterizing annual variations in transmission. Consider an outbreak occurring in year y. Given the initial state of the system, it is straightforward to simulate epidemics from the model (we use a discrete time-step of $\Delta T = 0.25$ days). The epidemic starts at time-step $L = 0$.

Denote by Z_L the history of the epidemic (specifying those in the population who are infected and those who are immune, and the times of infection) up to time-step L. Given Z_{L-1}, the probability that susceptible person i is infected during time-step L is:

$$p_{i,y}^{L} = 1 - \exp\{-\lambda_{i,y}(L\Delta T)\Delta T\}$$

where $\lambda_{i,y}(t)$ is the hazard of infection. The associated density is:

$$P(Z_L|Z_{L-1},\theta,V) = \prod_{i:\text{ susceptible people at } L-1} \left(p_{i,y}^{L}\right)^{x_i^L} \left(1-p_{i,y}^{L}\right)^{1-x_i^L}$$

where x_i^L is equal to 1 if susceptible person i is infected at time-step L and zero otherwise. The unconstrained density for the complete history of the epidemic Z is:

$$g(Z|\theta,V) = P(Z_0|Y,\theta) \prod_{L=1}^{\infty} P(Z_L|Z_{L-1},\theta,V) \quad (1)$$

Seeding the initial state of the system $\{Z_0\}$ is described in the Supplementary Information.

Simulation of constrained epidemics. We modify the unconstrained simulator to a constrained simulator, which simulates epidemics consistent with the data. Assume that the constrained epidemic has been simulated up to time-step $L-1$. From the model, we compute the risk of infection $\lambda_{i,y}(L\Delta T)$ for each susceptible person i for time-step L; and derive the expected incidence for adults $E_A(L)$ and children $E_C(L)$:

$$E_A(L) = \frac{\sum_{i:a(i)=A} \lambda_{i,y}(L\Delta T)\Delta T}{N_{com}} \text{ and } E_C(L) = \frac{\sum_{i:a(i)=C} \lambda_{i,y}(L\Delta T)\Delta T}{N_{com}}$$

We denote by $Y_A(L)$ and $Y_C(L)$ the observed incidence at time step L among adults and children respectively. The ratio of observed incidence to expected incidence is $\rho_A(L) = Y_A(L)/E_A(L)$ for adults and $\rho_C(L) = Y_C(L)/E_C(L)$ for children. The constrained epidemic is obtained by simulating the infection process with corrected infection risks. For any susceptible person i, the corrected infection risk is:

$$\lambda_{i,y}^{*}(L\Delta T) = \rho_{a(i)}(L)\lambda_{i,y}(L\Delta T)$$

where $a(i)$ is the age (either adult or child status) of the susceptible person. It is straightforward to check that, for the constrained process, the expected incidence at time step L is $Y_A(L)$ for adults and $Y_C(L)$ for children.

The density of the constrained simulation for time-step L is:

$$P_{\text{constrained}}(Z_L|Z_{L-1},Y,\theta,V) = \prod_{i:\text{ susceptible people at } L-1} \left(p_{i,y}^{L*}\right)^{x_i^L} \left(1-p_{i,y}^{L*}\right)^{1-x_i^L}$$

where $p_{i,y}^{L*} = 1 - \exp\left\{-\lambda_{i,y}^{*}(L\Delta T)\Delta T\right\}$ is the constrained probability of infection. The constrained density for the complete history of the epidemic Z is:

$$h(Z|Y,\theta,V) = P(Z_0|Y,\theta) \prod_{L=1}^{\infty} P_{\text{constrained}}(Z_L|Z_{L-1},Y,\theta,V) \quad (2)$$

Approximation of the likelihood. If the complete history of the epidemic Z (who has been infected when in the structured population) was known, it would be easy to write down the probability $P(Z|\theta,V)$ (see Supplementary Information), and then likelihood-based inference, in a frequentist or bayesian setting, would be straightforward. However, the data Y consist only of daily incidences. The likelihood $P(Y|\theta,V)$ is then difficult to compute because it requires integration with respect to the (unobserved) complete history Z, which has a very high dimension:

$$P(Y|\theta,V) = \int_Z P(Y|Z)g(Z|\theta,V)\,dZ$$

The first term of the integrand is the observation model, which ensures that complete history Z is consistent with the observed curves Y: so $P(Y|Z)$ is equal to 1 if Z is consistent with Y and 0 otherwise. The second term is the sampling density for the complete history Z of the epidemic (equation (1)).

We have designed an approach based on sequential importance sampling[20,21] to approximate the likelihood. The idea is to work with simulated epidemics that are constrained to be consistent with the observation Y (see above), and have density h (equation (2)). We can rewrite the likelihood (via multiplication by 1) as:

$$P(Y|\theta,V) = \int_Z \frac{g(Z|\theta,V)}{h(Z|Y,\theta_{\text{simul}},V_{\text{simul}})} h(Z|Y,\theta_{\text{simul}},V_{\text{simul}})dZ$$

To obtain an importance sampling approximation[21] of this integral, we simulate $Z^1,\ldots,Z^N$ constrained epidemics (sampling from density h, with parameters $\{\theta_{\text{simul}}, V_{\text{simul}}\}$), and approximate the likelihood by:

$$P(Y|\theta,V) \approx \frac{1}{N}\sum_{n=1}^{N} \frac{g(Z|\theta,V)}{h(Z|Y,\theta_{\text{simul}},V_{\text{simul}})}$$

There is much less stochastic fluctuation in the sampling process than for unconstrained epidemics, so we do not have to simulate large numbers of epidemics per observed curve. We found that the integral was well evaluated using only one simulation per observed curve ($N = 1$; see Supplementary Information). So, in practice, we used $N = 1$. This point makes the whole estimation process extremely efficient.

Another typical feature of sequential importance sampling is that the epidemic trajectory supporting estimation of the log-likelihood does not have to be obtained with the parameters of interest, that is $\{\theta_{\text{simul}}, V_{\text{simul}}\}$ may be different from $\{\theta, V\}$. Therefore, using a trajectory simulated with well-chosen values of $\{\theta_{\text{simul}}, V_{\text{simul}}\}$, the approximation of the log-likelihood is possible in a larger region of the parameter space. This property is at the heart of our estimates of the annual variations in transmission V (see Supplementary Information).

Statistical framework. In a bayesian context, we explore the posterior distribution of the parameters by Markov-chain Monte Carlo sampling[22]. When we account for annual variations in influenza transmission V, we rely on the profile likelihood for θ:

$$\text{LP}(\theta,Y) = P\left(Y|\theta,\hat{V}(\theta)\right)$$

where $\hat{V}(\theta)$ maximizes $P(Y|\theta,V)$ with respect to V and satisfies identifiability constraints (see Supplementary Information).

LETTERS

A plant pathogen virulence factor inhibits the eukaryotic proteasome by a novel mechanism

Michael Groll[1]*, Barbara Schellenberg[2]*, André S. Bachmann[3,4], Crystal R. Archer[3,4], Robert Huber[5,6,7], Tracy K. Powell[8], Steven Lindow[8], Markus Kaiser[9] & Robert Dudler[2]

Pathogenic bacteria often use effector molecules to increase virulence. In most cases, the mode of action of effectors remains unknown. Strains of *Pseudomonas syringae* pv. *syringae* (*Pss*) secrete syringolin A (SylA), a product of a mixed non-ribosomal peptide/polyketide synthetase, *in planta*[1]. Here we identify SylA as a virulence factor because a SylA-negative mutant in *Pss* strain B728a obtained by gene disruption was markedly less virulent on its host, *Phaseolus vulgaris* (bean). We show that SylA irreversibly inhibits all three catalytic activities of eukaryotic proteasomes, thus adding proteasome inhibition to the repertoire of modes of action of virulence factors. The crystal structure of the yeast proteasome in complex with SylA revealed a novel mechanism of covalent binding to the catalytic subunits. Thus, SylA defines a new class of proteasome inhibitors that includes glidobactin A (GlbA), a structurally related compound from an unknown species of the order Burkholderiales[2], for which we demonstrate a similar proteasome inhibition mechanism. As proteasome inhibitors are a promising class of anti-tumour agents, the discovery of a novel family of inhibitory natural products, which we refer to as syrbactins, may also have implications for the development of anti-cancer drugs[3]. Homologues of SylA and GlbA synthetase genes are found in some other pathogenic bacteria, including the human pathogen *Burkholderia pseudomallei*, the causative agent of melioidosis[4]. It is thus possible that these bacteria are capable of producing proteasome inhibitors of the syrbactin class.

Pss is a pathogen of many plant species, causing, for example, brown spot disease on bean (*P. vulgaris*)[5]. Some *Pss* strains secrete SylA, a peptide derivative synthesized by a mixed non-ribosomal peptide synthetase (NRPS)/polyketide synthetase (PKS) encoded by a gene cluster which was previously cloned from the strain *Pss* B301D-R[2]. To examine the functional role of SylA, we constructed a SylA-negative mutant by targeted gene disruption of the *sylC* gene in *Pss* B728a, a strain that causes brown spot disease on bean (*P. vulgaris*) and whose SylA synthetase gene cluster is almost identical to that of strain B301D-R[2]. The *sylC* gene encodes the first NRPS module of the SylA synthetase, and disruption of *sylC* in the B301D-R strain was shown to abolish SylA synthesis[2]. Infection experiments on bean plants revealed a significantly reduced virulence of the mutant strain. The number of brown spots developing on bean leaves after spray inoculation was reduced to 29 ± 21% compared with the wild type in four independent experiments (Fig. 1). Thus, SylA was identified as a virulence factor in this plant–pathogen interaction.

Previously performed transcript profiling revealed that exogenous application of SylA on wheat and *Arabidopsis thaliana* leads to pronounced changes in global gene activity[6]. Intriguingly, transcripts encoding all proteasome subunits and many heat-shock proteins accumulated in *Arabidopsis* leaves. Similar observations were reported for yeast and mammalian cells after treatment with proteasome inhibitors[7,8]. Therefore, we tested whether SylA was a proteasome inhibitor. The eukaryotic 20S proteasome contains three catalytic subunits (β1, β2 and β5) conferring caspase-like, trypsin-like and chymotrypsin-like proteolytic activities, respectively[3]. *In vitro* experiments showed that SylA inhibited all of them (Fig. 2a–d). In addition, measurement of proteasome activity in the presence of SylA revealed that the proteolytic reaction velocity was not constant but diminished as a function of time. This is similar to what is observed in the presence of the irreversible proteasome inhibitor epoxomicin and in contrast to the effect of the reversible peptide aldehyde inhibitor MG-132 (Fig. 2a), suggesting that SylA acts as an irreversible proteasome inhibitor. This finding was confirmed by removing inhibitors and substrate from chymotrypsin-like cleavage reactions by dialysis. Addition of substrate to dialysed proteasomes previously inhibited by MG-132 led to renewed activity, whereas proteasomes that had been inhibited by epoxomicin or SylA remained inactive (Supplementary Fig. 1). The chymotrypsin-like activity was found to be most sensitive to SylA inhibition, whereas higher concentrations of SylA were necessary to

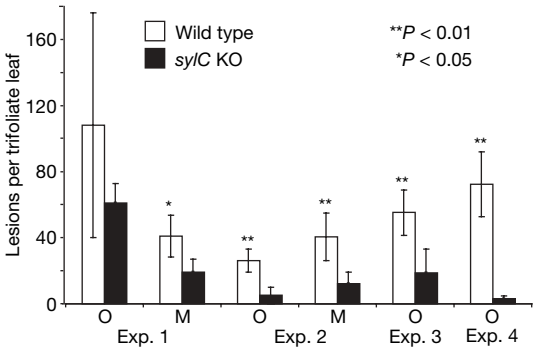


Figure 1 | Syringolin-negative mutant exhibits reduced virulence. Five pots per experiment (Exp), each with eight 18-day-old bean plants, were spray-inoculated with 10^5 cells per millilitre of wild-type or SylA-negative (*sylC* KO) strains of *Pss* B728a. Lesion numbers per trifoliate leaf were counted on the oldest (O) and middle-aged (M) leaves. Mean lesion numbers ± s.d. over the five replica pots are given. *p*, error probability (two-sided *t*-test).

[1]Center for Integrated Protein Science at the Department Chemie, Lehrstuhl für Biochemie, Technische Universität München, Lichtenbergstrasse 4, Garching D-85747, Germany. [2]Zurich–Basel Plant Science Center, Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland. [3]Cancer Research Center of Hawaii, University of Hawaii at Manoa, 1236 Lauhala Street, Honolulu, Hawaii 96813, USA. [4]Cell and Molecular Biology Graduate Program, John A. Burns School of Medicine, University of Hawaii at Manoa, 651 Ilalo Street, Honolulu, Hawaii 96813, USA. [5]Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany. [6]School of Biosciences, Cardiff University, Cardiff CF10 3US, UK. [7]Zentrum für medizinische Biotechnologie, Universität Duisburg-Essen, D-45117 Essen, Germany. [8]Department of Plant and Microbial Biology, University of California, 111 Koshland Hall, Berkeley, California 94720-3102, USA. [9]Chemical Genomics Centre of the Max Planck Society, Otto-Hahn-Strasse 15, D-44227 Dortmund, Germany.
*These authors contributed equally to this work.

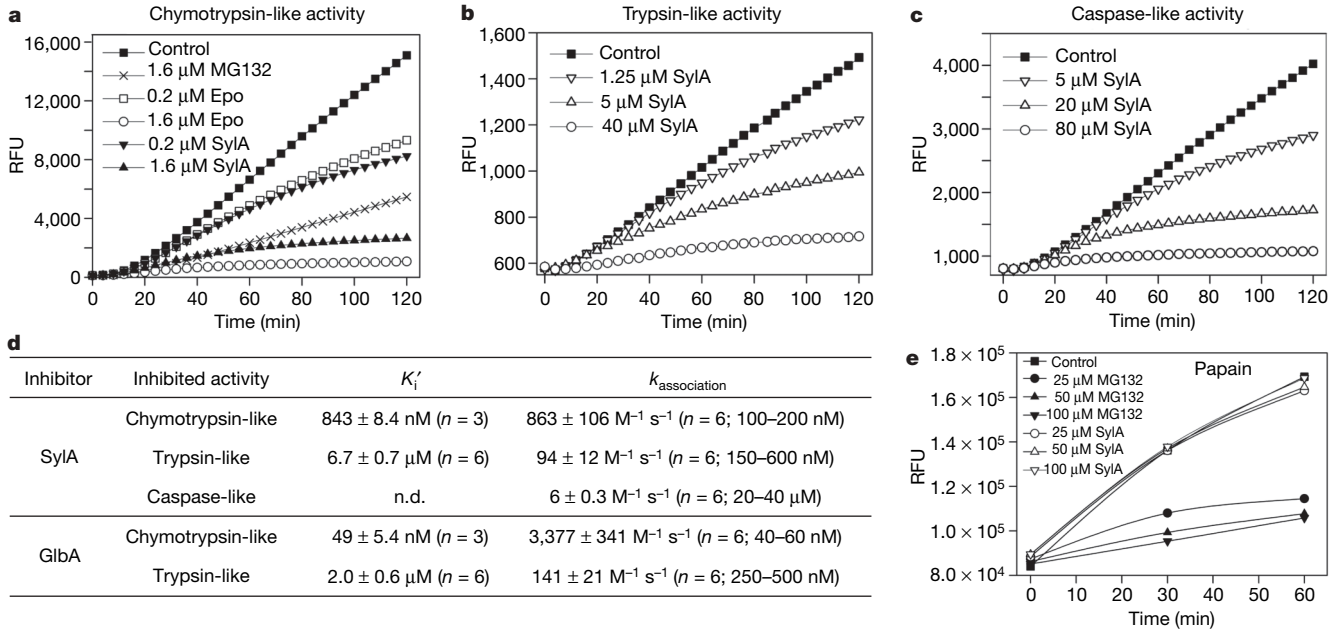
inhibit trypsin-like and caspase-like activities (Fig. 2b–d). In contrast to MG-132, which also inhibits cysteine proteases, SylA inhibition of the proteasome was specific because both papain (Fig. 2e) and the serine protease trypsin were not affected by SylA, even at significantly higher concentrations (up to 500 μM). SylA also inhibited the chymotrypsin-like activity of the proteasome in crude extracts of etiolated bean seedlings (Supplementary Fig. 2).

We next determined whether SylA inhibits *in vivo* proteasome function in *Arabidopsis* seedlings and human cultured cells. To this end, an *Arabidopsis* line homozygous for a *CyclinB1;1::uidA (GUS)* reporter fusion gene[9] was used. The mitotic *CyclinB1;1* gene is expressed only around the G2/M transition of the cell cycle[10]. The fusion gene contains a cyclin destruction box leading to the proteasomal degradation of the protein at the end of the M phase[11]. Proteasomal dysfunction due to a mutation has been shown to result in strong GUS staining of dividing cells[12]. Indeed, much stronger staining of dividing cells in the root meristem of 3-day-old seedlings treated with 10 μM SylA solution compared with controls was observed (Supplementary Fig. 3). Incubation of human SK-N-SH neuroblastoma cells with SylA for 2 h resulted in a dose-dependent decrease of proteasome activity (Supplementary Fig. 4a). As expected[13], incubation of SK-N-SH cells with 25 μM SylA resulted in a time-dependent accumulation of ubiquitinated proteins (Supplementary Fig. 4d). The SylA treatment of SK-N-SH cells was further accompanied by a time-dependent increase in the protein levels of the tumour suppressor protein p53, a known target of the proteasome and an important regulator of apoptosis and cell-cycle progression (Supplementary Fig. 4d).

To elucidate the binding mode of SylA to the proteasome we co-crystallized SylA with the yeast 20S proteasome and collected data to 2.9 Å resolution. Crystallographic refinement started from the coordinates of the yeast 20S proteasome[14] followed by anisotropic overall temperature factor correction and positional refinement using the Crystallography & NMR system (CNS) software suite (R_{crys}/R_{free} = 21.8/24.8; Supplementary Table 1)[15]. Electron-density maps calculated with phases after twofold averaging allowed a detailed interpretation of SylA (Fig. 3c), revealing that SylA covalently binds to the hydroxy group of the active site amino (N)-terminal threonine by a novel mechanism: Thr1O$^{\gamma}$ of the proteasome performs a Michael-type 1,4-addition to the double bond located at C4 in the 12-membered ring system of the inhibitor. The resulting covalent ether bond with the 12-membered ring system causes the irreversible inhibition. This reaction is facilitated by Gly47N, which stabilizes the carbonyl anion of SylA in its activated transition intermediate state by hydrogen bonds (oxyanion hole) (Fig. 3b). Furthermore, the 12-membered ring structure of SylA containing the functional reactive group has a constrained conformation favouring high-affinity binding for entropic reasons compared with flexible ligands (Fig. 3e). The functional reactive double bond at the 12-membered ring is combined with a dipeptide located in its close proximity (Fig. 3a). This dipeptide is essential for stabilization of SylA by formation of an antiparallel β-sheet at the substrate-binding channel, which increases the mean residence time of the inhibitors at the active centre for completing the covalent ether bond formation with Thr1O$^{\gamma}$.

The peptide structure of SylA is closely related to glidobactin A (GlbA; Fig. 3a), an acylated peptide derivative reported to have antifungal and anti-tumour activity[16]. Thus, we tested GlbA isolated from strain K481-B101 (ATCC 53080; DSM 7029)[4] for proteasome inhibition. Indeed, GlbA blocked the chymotrypsin-like activity irreversibly at low concentrations, whereas the trypsin-like activity was less sensitive (Fig. 2d) and the caspase-like activity was not inhibited at the concentrations tested (up to 20 μM). The crystal structure of GlbA in complex with the proteasome was elucidated at 2.7 Å resolution, yielding R_{crys}/R_{free} = 23.4/26.2 (Supplementary Table 1) (Fig. 3d). As expected, the complex structure revealed the same mechanism of inactivation of the chymotrypsin-like (Fig. 3f) and trypsin-like activities as SylA, whereas GlbA did not bind to the caspase-like active site.

The results described above show that SylA is a virulence factor that inhibits proteasome function, thus adding proteasome inhibition to the repertoire of modes of action of such factors. Although many virulence-related factors are delivered into host cells by type III secretion systems[17–19], SylA is thought to be secreted by an exporter belonging to the major facilitator superfamily[2]. This, together with the fact that exogenous application of SylA leads to proteasome inhibition (this report), inhibition of cell proliferation and induction of apoptosis[20] in mammalian cancer cells, as well as changes in gene activity typical for proteasome inhibition in *Arabidopsis* and wheat,



d

Inhibitor	Inhibited activity	K_i'	$k_{association}$
SylA	Chymotrypsin-like	843 ± 8.4 nM (*n* = 3)	863 ± 106 M^{-1} s^{-1} (*n* = 6; 100–200 nM)
	Trypsin-like	6.7 ± 0.7 μM (*n* = 6)	94 ± 12 M^{-1} s^{-1} (*n* = 6; 150–600 nM)
	Caspase-like	n.d.	6 ± 0.3 M^{-1} s^{-1} (*n* = 6; 20–40 μM)
GlbA	Chymotrypsin-like	49 ± 5.4 nM (*n* = 3)	3,377 ± 341 M^{-1} s^{-1} (*n* = 6; 40–60 nM)
	Trypsin-like	2.0 ± 0.6 μM (*n* = 6)	141 ± 21 M^{-1} s^{-1} (*n* = 6; 250–500 nM)

Figure 2 | SylA inhibits the eukaryotic proteasome. a–c, Inhibition of chymotrypsin-like **(a)**, trypsin-like (**b**) and caspase-like (**c**) activities of mammalian proteasomes by SylA, epoxomicin (Epo) and MG-132 using fluorogenic substrates. RFU, relative fluorescence units. **d**, Apparent K_i' values and rates of covalent inhibition ($k_{association}$) over inhibitor concentrations given in parentheses were derived from plots as shown in **a–c** (see Methods). Values represent means ± s.d. **e**, Cleavage of fluorescent casein by papain in the presence of MG-132 and SylA.

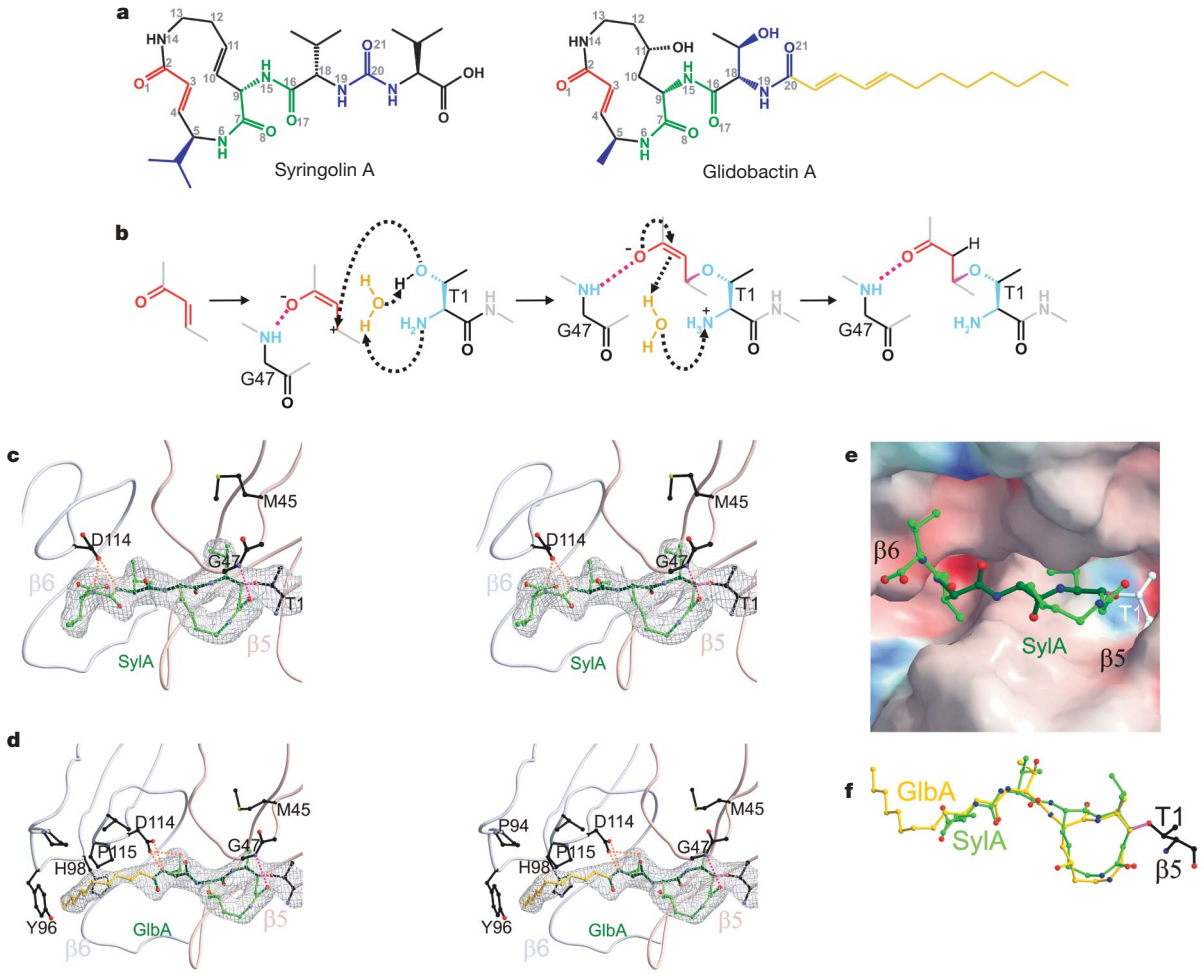


Figure 3 | Structural basis for proteasome inhibition by syrbactins. **a**, Chemical structure of SylA and GlbA. Red, α,β-unsaturated carbonyl group reacting with Thr1O$^{\gamma}$; green, dipeptide bond stabilizing the inhibitor upon proteasome binding; blue, molecule part determining active site specificity; yellow, aliphatic tail of GlbA. **b**, Mechanism of binding of SylA/GlbA to the active site Thr1. **c**, **d**, Stereo representation of the chymotrypsin-like active site (rose, subunit β5; light blue, subunit β6) in complex with (**c**) SylA (green; PDB accession code 2ZCY) and (**d**) GlbA (green, aliphatic tail in yellow; PDB accession code 3BDM). Magenta, covalent linkage of inhibitors with active site Thr1; dotted lines indicate hydrogen bonds. Black, residues performing specific interactions with SylA and GlbA. Electron-density maps (grey) are contoured from 1σ in similar orientations around Thr1. **e**, Electrostatic potential surface (contoured from +15kT/e (intense blue) to −15kT/e (intense red)) of SylA bound to subunit β5. **f**, Structural superposition of SylA (green) with GlbA (yellow) bound to subunit β5.

indicates that SylA is taken up by cells. As SylA is a hydrophilic molecule, it probably cannot passively cross membranes, and thus an uptake transporter must be postulated, which remains to be identified.

Previous studies have shown that the host proteasome has been instrumentalized by plant pathogens to suppress host defence reactions by type III effectors mediating proteasomal degradation of specific host proteins[2,21,22]. On the other hand, there is increasing evidence that the ubiquitin–proteasome degradation pathway is essential for pathogen defence and disease immunity of plants[23,24]. *Pss* appears to capitalize on this by inhibiting the host proteasome using SylA, thereby suppressing host defence reactions directly or indirectly. It also remains to be seen whether the proteasome is the sole target of SylA.

A unique model completely explaining the synthesis of the peptide part of SylA, including the functionally important double bond at C4 of the 12-membered ring system, has been proposed based on the architecture of the *sylC* and *sylD* genes in the SylA synthetase gene cluster[2]. The model allowed the recent cloning of genes responsible for GlbA synthesis, including the *sylC* and *sylD* homologues *glbF* and *glbC*, from K481-B101, the strain from which GlbA was originally isolated[16] and which belongs to an unknown species of the order Burkholderiales[4]. A National Center for Biotechnology Information database search of all sequenced bacterial genomes revealed homologues of these genes in the insect pathogen *Photorhabdus luminescens* and the human pathogen *B. pseudomallei*, the causative agent of melioidosis[4]. These organisms are therefore hypothesized to be capable of synthesizing proteasome inhibitors of the syrbactin class. If this prediction is confirmed, it will be interesting to determine whether the respective compounds are involved in virulence.

Proteasome inhibitors form a promising new therapeutic class of drugs against cancer and other diseases, causing intense interest in such molecules[3,25]. Indeed, SylA was recently shown to inhibit proliferation and induce apoptosis in neuroblastoma and ovarian cancer cells[20]. Elucidation of the structural inhibition motif of syrbactins opens new perspectives in the identification of natural products with proteasome inhibition potential, as well as the design and creation of new specific, selective and efficient proteasomal inhibitors. New inhibitors derived from such natural products can find their potential application in drug development.

METHODS SUMMARY

Mutant construction and plant inoculations. Construction of the SylA-negative mutant by plasmid insertion into the *sylC* gene of *Pss* B728a was as described for the corresponding mutant in strain *Pss* B301D-R[2]. Virulence of *Pss*

strains was assessed by counting the number of lesions on leaves of bean (*P. vulgaris*) plants sprayed with suspensions of 10^5 cells per millilitre (ref. 26).

Isolation of SylA and GlbA. SylA and GlbA have been isolated as described[2,4,27]. SylA solutions with defined concentrations were prepared by dissolving weighted amounts of SylA in water. Concentrations of GlbA solutions were determined by measuring the absorption at 261 nm in methanol using an extinction coefficient of 35,000 M^{-1} cm^{-1} (ref. 16). For proteasome inhibition assays, stock solutions of GlbA in DMSO were diluted with water.

Proteasome/protease inhibition assays. *In vitro* proteasome inhibition assays were performed in 96-well microtiter plates with human erythrocyte 20S proteasomes using the Assay Kit for Drug Discovery AK-740 (Biomol). *In vivo* proteasome inhibition in cultured mammalian cells was measured using the proteasome-Glo cell-based luminescent assay (Promega) in solid white 96-well microtiter cell culture plates. Bovine pancreas trypsin (Pierce) and *Papaya latex* papain (Sigma-Aldrich) inhibition assays were performed with the SensoLyte Green Protease Fluorometric Assay Kit (AnaSpec).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 October 2007; accepted 28 January 2008.

1. Wäspi, U., Blanc, D., Winkler, T., Ruedi, P. & Dudler, R. Syringolin, a novel peptide elicitor from *Pseudomonas syringae* pv. *syringae* that induces resistance to *Pyricularia oryzae* in rice. *Mol. Plant Microbe Interact.* **11,** 727–733 (1998).
2. Amrein, H. *et al.* Functional analysis of genes involved in the synthesis of syringolin A by *Pseudomonas syringae* pv. *syringae* B301D-R. *Mol. Plant Microbe Interact.* **17,** 90–97 (2004).
3. Borissenko, L. & Groll, M. 20S proteasome and its inhibitors: crystallographic knowledge for drug development. *Chem. Rev.* **107,** 687–717 (2007).
4. Schellenberg, B., Bigler, L. & Dudler, R. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. *Environ. Microbiol.* **9,** 1640–1650 (2007).
5. Hirano, S. S. & Upper, C. D. Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae* – a pathogen, ice nucleus, and epiphyte. *Microbiol. Mol. Biol. Rev.* **64,** 624–653 (2000).
6. Michel, K., Abderhalden, O., Bruggmann, R. & Dudler, R. Transcriptional changes in powdery mildew infected wheat and *Arabidopsis* leaves undergoing syringolin-triggered hypersensitive cell death at infection sites. *Plant Mol. Biol.* **62,** 561–578 (2006).
7. Fleming, J. A. *et al.* Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. *Proc. Natl Acad. Sci. USA* **99,** 1461–1466 (2002).
8. Meiners, S. *et al.* Inhibition of proteasome activity induces concerted expression of proteasome genes and *de novo* formation of mammalian proteasomes. *J. Biol. Chem.* **278,** 21517–21525 (2003).
9. Colon-Carmona, A., You, R., Haimovitch-Gal, T. & Doerner, P. Spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. *Plant J.* **20,** 503–508 (1999).
10. Doerner, P., Jorgensen, J. E., You, R., Steppuhn, J. & Lamb, C. Control of root growth and development by cyclin expression. *Nature* **380,** 520–523 (1996).
11. King, R. W., Deshaies, R. J., Peters, J. M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274,** 1652–1659 (1996).
12. Brukhin, V., Gheyselinck, J., Gagliardini, V., Genschik, P. & Grossniklaus, U. The RPN1 subunit of the 26S proteasome in *Arabidopsis* is essential for embryogenesis. *Plant Cell* **17,** 2723–2737 (2005).
13. Meng, L. H. *et al.* Epoxomicin, a potent and selective proteasome inhibitor, exhibits *in vivo* antiinflammatory activity. *Proc. Natl Acad. Sci. USA* **96,** 10403–10408 (1999).
14. Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386,** 463–471 (1997).
15. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
16. Oka, M. *et al.* Glidobactins A, B and C, new antitumor antibiotics. I. Production, isolation, chemical properties and biological activity. *J. Antibiot. (Tokyo)* **41,** 1331–1337 (1988).
17. Desveaux, D., Singer, A. U. & Dangl, J. L. Type III effector proteins: doppelgangers of bacterial virulence. *Curr. Opin. Plant Biol.* **9,** 376–382 (2006).
18. Galán, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444,** 567–573 (2006).
19. Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M. & Dangl, J. L. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* **60,** 425–449 (2006).
20. Coleman, C. S. *et al.* Syringolin A, a new plant elicitor from the phytopathogenic bacterium *Pseudomonas syringae* pv. *syringae*, inhibits the proliferation of neuroblastoma and ovarian cancer cells and induces apoptosis. *Cell Prolif.* **39,** 599–609 (2006).
21. Nomura, K. *et al.* A bacterial virulence protein suppresses host innate immunity to cause plant disease. *Science* **313,** 220–223 (2006).
22. Rosebrock, T. R. *et al.* A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. *Nature* **448,** 370–374 (2007).
23. Dong, W. B., Nowara, D. & Schweizer, P. Protein polyubiquitination plays a role in basal host resistance of barley. *Plant Cell* **18,** 3321–3331 (2006).
24. Goritschnig, S., Zhang, Y. L. & Li, X. The ubiquitin pathway is required for innate immunity in *Arabidopsis*. *Plant J.* **49,** 540–551 (2007).
25. Adams, J. Proteasome inhibitors as new anticancer drugs. *Curr. Opin. Oncol.* **14,** 628–634 (2002).
26. Quinones, B., Dulla, G. & Lindow, S. E. Quorum sensing regulates exopolysaccharide production, motility, and virulence in *Pseudomonas syringae*. *Mol. Plant Microbe Interact.* **18,** 682–693 (2005).
27. Wäspi, U., Schweizer, P. & Dudler, R. Syringolin reprograms wheat to undergo hypersensitive cell death in a compatible interaction with powdery mildew. *Plant Cell* **13,** 153–161 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Grossniklaus and J. Celenza (Boston University) for the *CycB1;1-GUS* line. U. Grossniklaus and H. Gehring are thanked for reading the manuscript. We also thank D. Albert, H. Gehring, Z. Hasenkamp, R. Go, D. Koomoa, J. Molnar, A. Niewienda and C. Wallick for technical advice and assistance. We are grateful to L. Eberl for use of equipment. C.R.A was supported by a graduate student research assistantship from the Cell and Molecular Biology Graduate Program, University of Hawaii. Support by grants from the Swiss National Science Foundation to R.D. is acknowledged.

Author Contributions M.K., M.G., S.L., R.D. and A.S.B. designed experiments, analysed results and wrote the manuscript. R.D. constructed the sylA-negative mutant, T.K.P. performed the bean inoculation experiments, B.S. isolated SylA and GlbA, and performed *in vitro* proteasome and protease inhibition assays. M.G., M.K and R.H. performed and analysed crystallization experiments. C.R.A. performed the mammalian cell-based proteasome inhibition assays and immunoblot studies.

Author Information Atomic coordinates have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org/pdb) under the accession code 2ZCY (yeast 20S proteasome–syringolin A complex) and 3BDM (yeast 20S proteasome–glidobactin A complex). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.D. (rdudler@botinst.uzh.ch), A.S.B. (abachmann@crch.hawaii.edu) or M.K. (markus.kaiser@cgc.mpg.de).

METHODS

***In vitro* proteasome/protease assays and determination of constants.** Proteasome inhibition assays with human erythrocyte 20S proteasomes (AK-740, Biomol) were performed at 37 °C in 100 µl reaction volumes containing 2 µg ml^{-1} 20S proteasome and 100 µM Suc-LLVY-AMC, Boc-LRR-AMC or Z-LLE-AMC for assaying the chymotrypsin-like, trypsin-like and caspase-like activity, respectively. Fluorescence was monitored with an MWGt Sirius HT plate reader (BIO-TEK Instruments) equipped with 360 nm excitation and 460 nm emission filters.

Fluorescence values (RFU) measured in proteasome inhibition assays were plotted versus time (t). The fluorescence data (see Fig. 2) were fitted by the least squares method using SigmaPlot version 10.0 software (Systat Software) to the equation $f = f_0 + v_s t + [(v_i - v_s)/k_{obs}][1 - \exp(-k_{obs}t)]$, where v_i and v_s are initial and final velocities, respectively, and k_{obs} is the pseudo-first-order association rate constant[13,28]. Fittings with R^2 values > 0.99 were usually obtained. The rate of covalent inhibition ($k_{association}$) which equals $k_{obs}/[I]$, was calculated over the range of inhibitor concentrations given in parentheses in Fig. 2d. The initial velocities (v_i) were determined from the time derivative of these curves and plotted against the corresponding inhibitor concentration ([I]). Where possible, apparent K_i' values were determined by plotting v_i against the inhibitor concentration ([I]). Using SigmaPlot, the data were fitted to the hyperbolic equation $v_i = c + [v_0/(K_i' + [I])]$, where v_0 is the velocity of the control. Usually, R^2 values > 0.98 were obtained.

Bovine pancreas trypsin (Pierce) and *Papaya latex* papain (Sigma-Aldrich) inhibition assays were performed with the SensoLyteTM Green Protease Fluorometric Assay Kit (AnaSpec) in 100 µl volumes containing 200 ng ml^{-1} enzyme at 33 °C. For detection, 485/550 nm excitation/emission filters were used.

Inhibition assays of plant proteasomes were performed with crude lysates of 7-day-old etiolated bean seedlings. Hypocotyls and emerging first leaves of seedlings were ground in liquid nitrogen and extracted with twice the amount (v/w) of extraction buffer (50 mM Tris-HCl, pH 7.5, 2 mM EDTA, 100 mM NaCl, 10 mM sodium orthovanadates). After centrifugation at 500g for 15 min, the supernatant was used for inhibition experiments of the proteasome chymotrypsin-like activity in microtitre plates. Each reaction contained 20 µl lysate (about 100 µg protein), 100 µM Suc-LLVY-AMC substrate, desired concentrations of inhibitors and assay buffer (AK-740, Biomol) to a volume of 100 µl. In some experiments, the cysteine protease inhibitor E64 (Sigma) was included in each reaction at a final concentration of 50 µM. Reactions were monitored at room temperature every 4 min as described above. Reaction rates were approximately linear between the 8 and 32 min time points.

***In vivo* proteasome inhibition and immunoblot analysis.** Source and maintenance of the human neuroblastoma cell line SK-N-SH have been described[20,29]. For *in vivo* proteasome inhibition assays, wells of solid white 96-well microtiter cell culture plates were seeded with 1.1×10^5 cells per millilitre (90 µl per well). After 24 h, 10 µl SylA was added at different concentrations (0–100 µM) and incubated for 2 h. The proteasome inhibitor epoxomicin (0–1 µM) and the apoptosis-inducing agent cerulenin (0–50 µg ml^{-1})[30,31] were used as positive and negative controls, respectively (Supplementary Fig. 4). Microtiter plates were then equilibrated to room temperature for 15 min before 100 µl of proteasome Glo reagent (containing the bioluminescent substrate Suc-LLVY-aminoluciferin) per well was added and luminescence measured according to the manufacturer's instructions (Promega).

For immunoblot analysis, cells were seeded in six-well cell culture plates at a concentration of 3.8×10^5 cells per millilitre (1.98 ml per well). After 24 h, 20 µl of SylA (final concentration 25 µM) or cell culture medium (control, C) was added. For time-course experiments, cell treatments were staggered to process all samples at the same time after 24 h. Cell lysates were prepared as previously reported[29]. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE), electro-transfer to PVDF Immobilon-P membranes (Millipore) and antibody incubations were performed according to standard procedures. The primary antibodies used were rabbit whole serum ubiquitin (1:150) (U5379; Sigma), mouse monoclonal tumour suppressor protein p53 (1:250) (sc-126; Santa Cruz Biotechnology) and rabbit monoclonal α-tubulin (1:1,500) (11H10; Cell Signalling Technology). Horseradish peroxidase (HRP)-conjugated anti-mouse or anti-rabbit antibodies (1:5,000) were used as secondary antibodies. Membranes were developed using the ECL Plus kit following the manufacturer's protocol (Amersham Biosciences) and exposed to Blue Lite Autorad Film (ISC BioExpress). Membranes were stripped with ECL stripping buffer (62.5 mM Tris-HCl, pH 6.7, 2% SDS, 100 mM β-mercaptoethanol) for 20 min at 55 °C and sequentially re-probed with the next antibody. The experiment was repeated three times with similar results, as shown in Supplementary Fig. 4d.

Treatment and GUS-staining of transgenic *Arabidopsis* carrying a *CyclinB1;1::uidA* fusion gene. A homozygous line containing a transgene derived from pCDG in which the *CycB1;1* promoter plus the sequence encoding the first 150 amino acids of the CYCB1;1 protein including a cyclin destruction box were fused in frame to the *uidA* reporter gene encoding β-glucuronidase (GUS)[9]. Seeds of this line, which were provided by U. Grossniklaus, University of Zurich, were germinated on vertically oriented MS plates (4.4 g l^{-1} Murashige and Skoog basal medium with Gamborg's vitamins (Sigma), 3% sucrose and 0.6% Phytagel (Sigma)) that were incubated under a 16 h light, 8 h dark regime at 22 °C. Five-microlitre droplets of water with or without 10 µM SylA were applied to 3-day-old seedlings. After an 18 h incubation period, seedlings were transferred to a 96-well microtiter plate, stained for GUS activity for 60 min as described[32] and observed under a Leica DMR microscope.

Co-crystallization. Crystals of the 20S proteasome from *S. cerevisiae* were grown in hanging drops at 24 °C as described[14,33] and incubated for 60 min with SylA or GlbA. The protein concentration used for crystallization was 40 mg ml^{-1} in Tris-HCl (10 mM, pH 7.5) and EDTA (1 mM). The drops contained 3 µl of protein and 2 µl of the reservoir solution (30 mM magnesium acetate, 100 mM morpholino-ethane-sulphonic acid (pH 7.2) and 10% MPD).

The space group belongs to $P2_1$ with cell dimensions of about $a = 134$ Å, $b = 302$ Å, $c = 144$ Å and $\beta = 112°$ (see Supplementary Table 1). Data to 2.9 Å for the yeast 20S proteasome–SylA- and to 2.7 Å for the 20S proteasome–GlbA complexes were collected using synchrotron radiation with $\lambda = 1.05$ Å on the BW6-beamline at DESY, Hamburg, Germany, and with $\lambda = 1.0$ Å on the X06SA-beamline at SLS, Villingen, Switzerland, respectively. Crystals were soaked in a cryoprotecting buffer (30% MPD, 20 mM magnesium acetate, 100 mM morpholino-ethane-sulphonic acid, pH 6.9) and frozen in a stream of liquid nitrogen gas at 90 K (Oxford Cryo Systems). X-ray intensities were evaluated by using the DENZO program package[34] and data reduction was performed with CCP4[35]. The anisotropy of diffraction was corrected by an overall anisotropic temperature factor by comparing observed and calculated structure amplitudes using the program CNS[15]. Electron density was improved by averaging and back transforming the reflections ten times over the twofold non-crystallographic symmetry axis using the program package MAIN[36]. Conventional crystallographic rigid body, positional and temperature factor refinements were performed with CNS[15] using the yeast 20S proteasome structure as starting model[14]. Model building was performed using the program MAIN. Apart from the bound inhibitor molecules, structural changes were only noted in the specificity pockets. Temperature factor refinement indicates full occupancies of all inhibitor-binding sites. The inhibitors have been omitted for phasing.

28. Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268,** 726–731 (1995).
29. Wallick, C. J. *et al.* Key role for p27(Kip1), retinoblastoma protein Rb, and MYCN in polyamine inhibitor-induced G(1) cell cycle arrest in MYCN-amplified human neuroblastoma cells. *Oncogene* **24,** 5606–5618 (2005).
30. Geerts, D. *et al.* Expression of PRA1 domain family, member 2 (PRAF2) in neuroblastoma: correlation with clinical features, cellular localization, and cerulenin-mediated apoptosis regulation. *Clin. Cancer Res.* **13,** 6312–6319 (2007).
31. Heiligtag, S. J., Bredehorst, R. & David, K. A. Key role of mitochondria in cerulenin-mediated apoptosis. *Cell Death Differ.* **9,** 1017–1025 (2002).
32. Rodrigues-Pousada, R. A. *et al.* The *Arabidopsis* 1-aminocyclopropane-1-carboxylate synthase gene-1 is expressed during early development. *Plant Cell* **5,** 897–911 (1993).
33. Groll, M. & Huber, R. Purification, crystallization and X-ray analysis of the yeast 20S proteasomes. *Methods Enzymol.* **398,** 329–336 (2005).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
35. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59,** 1131–1137 (2003).
36. Turk, D. *Improvement of a Programme for Molecular Graphics and Manipulation of Electron Densities and Its Application for Protein Structure Determination.* PhD thesis, Technische Univ. München (1992).

LETTERS

Endothelins are vascular-derived axonal guidance cues for developing sympathetic neurons

Takako Makita[1], Henry M. Sucov[2], Cheryl E. Gariepy[3], Masashi Yanagisawa[4] & David D. Ginty[1]

During development, sympathetic neurons extend axons along a myriad of distinct trajectories, often consisting of arteries, to innervate one of a large variety of distinct final target tissues. Whether or not subsets of neurons within complex sympathetic ganglia are predetermined to innervate select end-organs is unknown. Here we demonstrate in mouse embryos that the endothelin family member Edn3 (ref. 1), acting through the endothelin receptor EdnrA (refs 2, 3), directs extension of axons of a subset of sympathetic neurons from the superior cervical ganglion to a preferred intermediate target, the external carotid artery, which serves as the gateway to select targets, including the salivary glands. These findings establish a previously unknown mechanism of axonal pathfinding involving vascular-derived endothelins, and have broad implications for endothelins as general mediators of axonal growth and guidance in the developing nervous system. Moreover, they suggest a model in which newborn sympathetic neurons distinguish and choose between distinct vascular trajectories to innervate their appropriate end organs.

The sympathetic division of the autonomic nervous system is composed of preganglionic sympathetic neurons, derived from the neuroectoderm, and their synaptic targets, neural-crest-derived postganglionic sympathetic neurons, the cell bodies of which reside within paravertebral and prevertebral sympathetic ganglia. Postganglionic sympathetic neurons (referred to hereafter as sympathetic neurons) innervate and control a wide range of targets, including exocrine and endocrine glands and ducts, smooth muscle layers of the intestine and blood vessels, cardiac muscle and nodes, and pilomotor muscles. Sympathetic neurons originate during development when neural crest cells bilaterally delaminate from the neural tube, migrate and assemble into ganglia. In mouse embryos, aggregation of neural crest cells into sympathetic chain ganglia initiates at approximately embryonic day 10.5 (~E10.5), adjacent to the dorsal aorta. The cranial-most sympathetic ganglia, the superior cervical ganglia (SCG), ultimately come to reside adjacent to the bifurcation of the internal and external carotid arteries from the common carotid arteries. Sympathetic neurons of the SCG send axonal projections either along the external carotid arteries to innervate the submandibular, sublingual and parotid glands or along the internal carotid arteries to the lacrimal and pineal glands, the eye, blood vessels and skin of the head, and the mucosa of the oral and nasal cavities.

A fundamental, unanswered question is how subsets of neurons within any complex autonomic ganglion, such as the SCG, acquire characteristic molecular and functional properties appropriate for the control of their respective end organs. One model posits that axons emanating from the SCG or other ganglia grow along intermediate targets, such as blood vessels, in a random or stochastic manner, and that it is the end organ that retrogradely specifies morphological and biochemical features, presynaptic connections and physiologic properties of the innervating neuron. Evidence in support of this model derives from the observation that a small subset of sympathetic neurons acquires a cholinergic transmitter phenotype on innervation of eccrine sweat glands and acquisition of a sweat-gland-derived cholinergic differentiation factor (CDF)[4]. A second model posits that subpopulations of newly born sympathetic neurons are molecularly distinct; these subpopulations choose from several intermediate target structures, such as particular blood vessels that serve to guide axonal trajectories to their appropriate end organs. Identification of axonal guidance cues that direct select subsets of sympathetic axons towards one or another vascular path would serve as evidence in support of the latter model. We therefore asked whether axons of developing sympathetic neurons of the SCG choose between extending along either the internal or the external carotid arteries because these blood vessels provide the main routes by which SCG axons project to distinct end-organs in the head and neck. These investigations led us to define the molecular identity of the cues and their receptors that guide extending axons along the vascular network.

In mouse embryos, the SCG form in the upper thorax by E11.5 as a condensation of neural precursors dorsal to the outflow tract of the heart. SCG neurons extend axons cranially along the primitive internal carotid arteries (derived from the dorsal aortae; Fig. 1a, d). At this stage, the external carotid arteries are beginning to emerge from the third pharyngeal arch arteries near their origins from the aortic sac[5]; there are no projections of sympathetic axons along either the nascent external carotid arteries or the third arch arteries (Fig. 1a, g). By E12.5, the common carotid arteries have elongated and the external carotid arteries are substantially developed. At this time, the sympathetic chain has become delimited into distinct ganglia and the SCG are clearly distinct from the more caudal stellate ganglia. Sympathetic neurons initiate extension of axons towards the external carotid arteries at this time, originating from the SCG at the level of the junction of the base of internal and external carotid arteries (Fig. 1b, m, p); the SCG extends cranially along the internal carotid arteries as far as the stapedial arteries (Fig. 1b, j). By E13.5, SCG axonal projections associated with both the external and internal carotid arteries are more substantially developed, although the general pattern of blood vessels and axonal projections is the same as the previous day (Fig. 1c, q, t, w and Supplementary Figs 1 and 2).

Previous studies using Wnt1Cre and R26R mice have defined the neural crest and mesodermal origins of vascular smooth muscle in the cardiac outflow tract and great blood vessels, although not specifically of the vascular segments relevant for SCG projections[6,7]. The

[1]The Solomon H. Snyder Department of Neuroscience, Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [2]Institute for Genetic Medicine, University of Southern California Keck School of Medicine, Los Angeles, California 90033, USA. [3]Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, Michigan 48109, USA. [4]Department of Molecular Genetics, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA.

third pharyngeal arch arteries are lined by neural-crest-derived smooth muscle (Fig. 1h, i). Consequently, their derivatives, which include the entirety of the external carotid arteries and the base of the internal carotid arteries, also have a neural crest origin (Fig. 1n, o, u, v and Supplementary Fig. 2). In contrast, the dorsal aortae are lined by smooth muscle that is mesodermal in origin (Fig. 1h, i) and, as a result, the internal carotid arteries from just above their origination from the common carotid arteries have a mesoderm-derived smooth muscle layer (Fig. 1e, f, k, l, r, s and Supplementary Fig. 2). Thus, the smooth muscle layers of the internal and external carotid arteries arise from different cell lineages and may therefore also be molecularly distinct. These arteries may, therefore, serve as distinct axonal growth environments that preferentially attract one, but not the other, of the two main axon bundles emanating from the SCG.

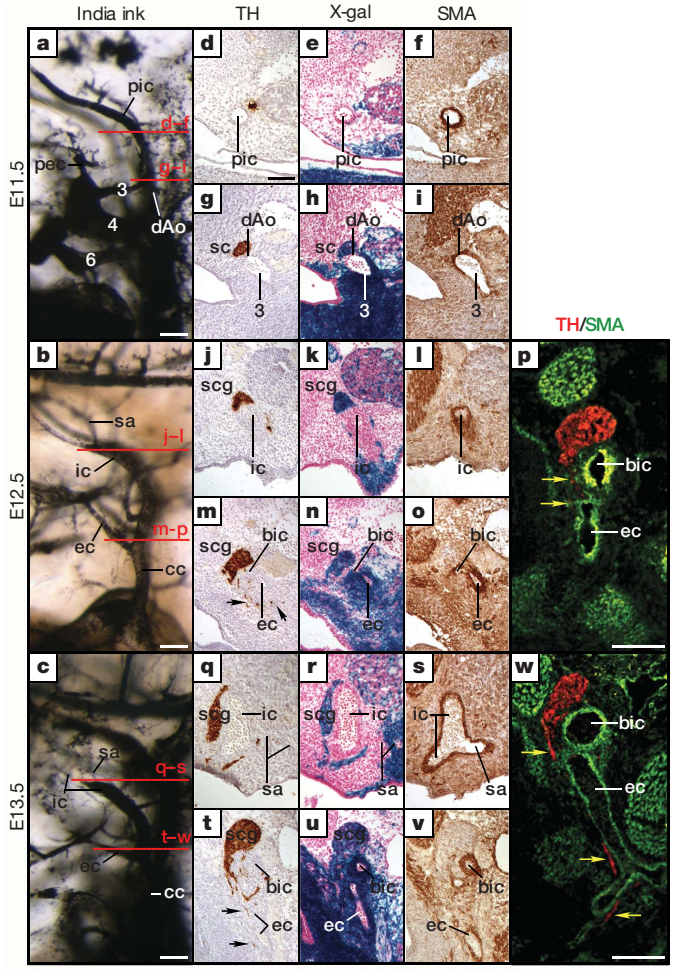


Figure 1 | Development of vascular and sympathetic systems. **a–c**, India ink visualization of the carotid arteries. The red lines indicate the approximate levels of sections shown in **d–w**. **d–w**, Serial transverse sections at the approximate levels indicated in panels **a–c** were stained with X-gal or immunostained for TH or alpha smooth muscle actin. **d–f**, **j–l** and **q–s** are sections through the internal carotid artery. **g–i**, **m–p** and **t–w** are at the level of the carotid bifurcation and show the base of the internal and external carotid arteries. **m**, **t**, Arrows denote SCG axons extending along the external carotid artery. Low-magnification views of the same sections are provided in Supplementary Fig. 1. At least ten embryos at each stage were analysed in this manner with comparable results. **p**, **w**, Confocal immunofluorescent images of sections of E12.5 and E13.5 embryos at a level approximately as indicated in **b** and **c**, respectively, showing TH-positive axons (red) projecting along the smooth muscle layer (green) of the base of internal and external carotid arteries (yellow arrows). Abbreviations: 3, 4 and 6, third, fourth and sixth pharyngeal arch arteries, respectively; bic, base of internal carotid artery; cc, common carotid artery; dAo, dorsal aorta; ec, external carotid artery; ic, internal carotid artery; pec, primitive external carotid artery; pic, primitive internal carotid artery; sa, stapedial artery; sc, sympathetic chain; and scg, superior cervical ganglion. Scale bars, 100 μm.

To address this possibility and to identify previously unknown vascular-derived guidance cues for SCG neurons that are selectively expressed by the crest-derived external carotid and the mesoderm-derived internal carotid arteries, we isolated these vascular segments from embryos at E13.5 and extracted RNA to screen microarrays for differentially expressed genes. Transcripts for endothelin converting enzyme 1 (ECE1), the protease responsible for conversion of inactive precursors (big endothelins) to bioactive forms of endothelins[8,9], were identified as being selectively expressed in the developing external carotid arteries.

To confirm this finding, and to address the role of endothelins and their cell surface receptors in SCG axonal guidance, we next examined the expression patterns of endothelin signalling components by *in situ* hybridization on serial sections from Wnt1Cre/R26R embryos at E11.5–E14.5. X-gal staining was used to identify neural-crest-derived cells, and alpha-smooth muscle actin (SMA) immunohistochemistry was used to label smooth muscle. We found that *Ece1* was selectively expressed in the neural-crest-derived smooth muscle layer of the external carotid arteries, but was absent, or present at much lower levels, in mesoderm-derived smooth muscle of the internal carotid arteries (Fig. 2b, h). *Ece1* mRNA was most abundant at E11.5 in the third pharyngeal arch arteries (the source of the external carotid arteries), was somewhat lower at E12.5 in the newly formed external carotid arteries, and was only marginally detectable at E13.5 and E14.5 (data not shown). At all stages examined, ECE1 was preferentially associated with the external carotid arteries relative to the internal carotid arteries. Expression of *Ece2* has been described, and is not found in this region until E13.5 (when it begins to be expressed in scattered clusters of neurons in the neural tube, dorsal root ganglia and sympathetic ganglia[10]). Of the three endothelins[1], endothelin 3 (*Edn3*) mRNA was found at the highest level, and was preferentially expressed around the third arch arteries at E11.5 and associated with the external carotid arteries at E12.5, although there was a small amount of expression detected around the internal carotid arteries at E11.5 and E12.5 (Fig. 2d, j). Similar to *Ece1*, *Edn3* expression was highest at E11.5, was somewhat diminished at E12.5, and was maintained at a low level at E13.5–E14.5. *Edn2* was not appreciably expressed in this region during this developmental period (Fig. 2e, k), and *Edn1* was expressed in endothelial cells and some scattered smooth muscle and mesenchymal cells at a relatively low level (Fig. 2f, l). Quantitative real-time PCR measurements using dissected internal and external carotid arteries confirmed the selective expression of *Ece1* and *Edn3* associated with the external carotid arteries (Fig. 2m). Both endothelin receptor genes, *Ednra* and *Ednrb*, encoding EdnrA and EdnrB, respectively[2,3], were expressed in subpopulations of SCG neurons (Fig. 2n–v). *Ednra* expression persists through E12.5, although expression abated by E13.5 (Fig. 2u) and was nearly undetectable at E14.5 (data not shown). Expression of *Ednrb*, in contrast, was sustained in a small subset of SCG neurons at E13.5 (Fig. 2v) and E14.5 (data not shown). Thus, although previously not appreciated to function in axonal growth and guidance, the spatial and temporal patterns of expression of endothelins, the ECE1 processing enzyme and the endothelin receptors suggest a role for endothelin signalling in guiding subsets of sympathetic axons towards either the internal or the external carotid arteries.

We next used cultured SCG explants grown in a collagen gel support[11] to ask whether endothelins have the capacity to influence axonal outgrowth of sympathetic axons *in vitro*. Indeed, robust extension of neurites from E12.5 SCG explants was observed when any of three endothelins was mixed uniformly into the collagen gel, whereas little or no neurite outgrowth was observed when bovine serum albumin (BSA) was provided. SCG explants obtained from E13.5 embryos were equally responsive to endothelins, but no response was observed in E14.5 SCG (Fig. 3a, Supplementary Fig. 3a). To ask whether endothelins are capable of attracting sympathetic

axons, we next placed E12.5 SCG explants in proximity to beads soaked with endothelins. Human EDN3 seemed to be most effective in this regard (Fig. 3b, middle, and Supplementary Fig. 3b), although EDN1 and EDN2 also promoted directional outgrowth (data not shown). Moreover, when beads soaked with a high concentration of EDN3 were placed adjacent to SCG explants, random or undirected outgrowth was observed in addition to directional outgrowth (Fig. 3b, right). Thus, axonal outgrowth is a concentration-dependent and possibly gradient-dependent process. To determine which endothelin receptor mediates this response, we treated E12.5–E13.5 SCGs with EDN3 in the presence or absence of endothelin-receptor-selective antagonists. The EdnrA antagonist BQ123 (ref. 12) completely eliminated SCG neurite outgrowth, whereas the EdnrB antagonist BQ788 (ref. 13) was ineffective (Fig. 3c, upper row, and Supplementary Fig. 3c). In contrast, outgrowth from the Xth cranial (CN X/vagal) ganglia explants, which express both endothelin receptors (data not shown), was selectively blocked by the EdnrB antagonist (Fig. 3c, lower row, and Supplementary Fig. 3d). We conclude that EDN3–EdnrA signalling is a potent mediator of SCG neurite outgrowth *in vitro*.

We next asked whether endothelins govern axonal guidance decisions of sympathetic neurons as they encounter the choice of projecting to either the internal or the external carotid arteries *in vivo*. For these experiments, SCG axonal projections in mice harbouring null mutations in genes encoding endothelin signalling components were assessed. Although EDN2 can promote axonal extension *in vitro*, *Edn2* is not expressed in vascular smooth muscle, and *Edn2*-null mouse embryos exhibited normal SCG axonal growth along both the internal and the external carotid arteries (data not shown). Similarly, *spotting lethal* mutant rat embryos, which carry a loss-of-function mutation in the *Ednrb* gene[14], also displayed normal SCG axonal projections (Supplementary Fig. 4). Thus, consistent with our

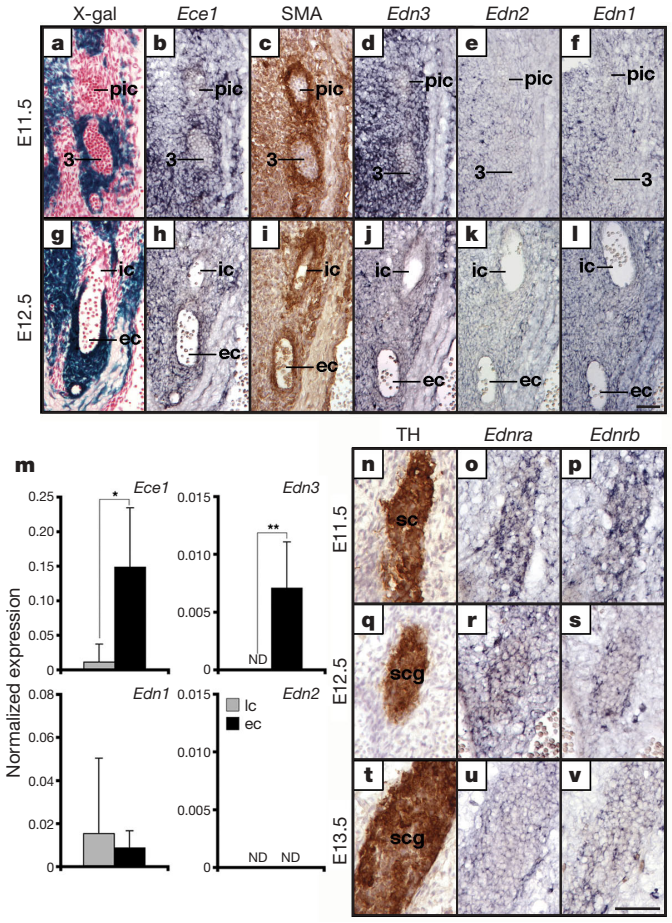


Figure 2 | Expression of endothelin signalling components. a–l, Serial frontal sections of E11.5 (**a–f**) and E12.5 (**g–l**) Wnt1Cre/R26R embryos, stained with X-gal, immunostained for SMA, or digoxygenin-labelled by *in situ* hybridization for the indicated genes. **m**, Quantitative real-time PCR analysis. Total RNA was extracted from dissected vascular segments from five individual E12.5 wild-type embryos isolated from two different litters. Total RNA was reverse transcribed, and was evaluated for gene expression by quantitative real-time PCR as described in the Methods. Expression levels were normalized to the amount of GAPDH transcripts and are scaled as absolute values. ic and ec, internal and external carotid artery segments. The grey bars represent the signal associated with the internal carotid arteries, whereas the black bars represent the signal associated with the external carotid arteries. ND, not detected. Error bars, mean ± s.e.m.; $*P < 0.01$; $**P < 0.005$. **n–v**, Serial frontal sections of embryos at the indicated stages immunolabelled for TH or labelled by *in situ* hybridization for *Ednra* and *Ednrb*. Sections shown at E11.5 are through the rostral end of the sympathetic chain, and at E12.5 and E13.5 are through the SCG. Results were confirmed with sections from at least five embryos of each stage. Scale bars, 100 μm.

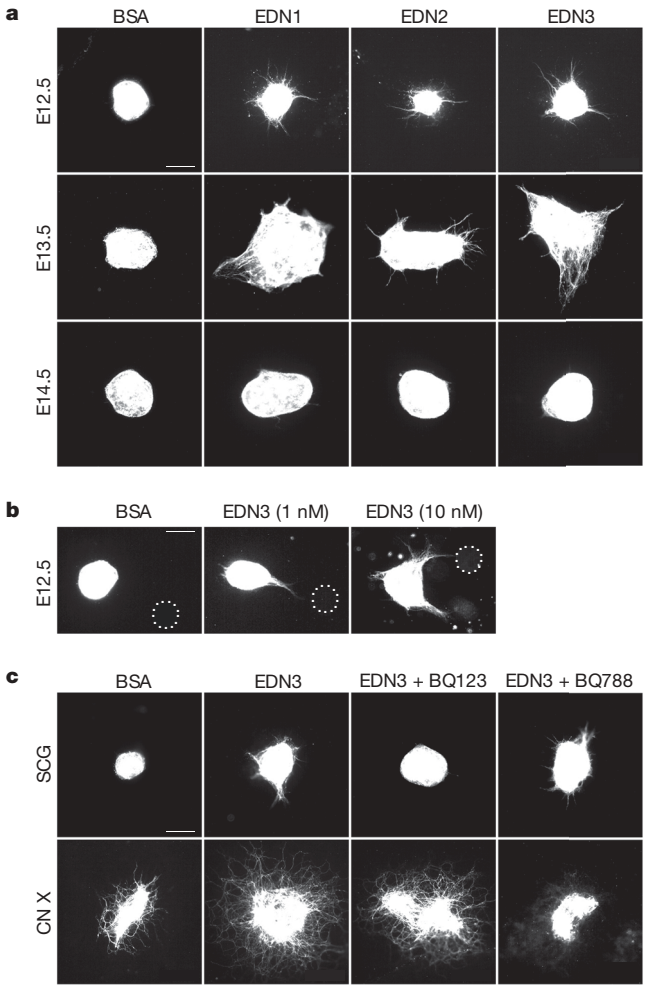


Figure 3 | Endothelins promote SCG neurite outgrowth *in vitro*. a, SCGs from embryos at the indicated stages were harvested and grown in collagen gels prepared with the indicated endothelins (10 nM) or BSA. The average amount of neurite outgrowth from seven separate experiments is shown in Supplementary Fig. 3a. **b**, Heparin–agarose beads soaked in the indicated concentration of EDN3 or BSA were placed adjacent to E12.5 SCG. Dotted circles indicate the location of the beads. The extent of directional outgrowth was assessed by the relative amount of outgrowth in the proximal direction relative to the distal direction of the beads, as shown in Supplementary Fig. 3b. **c**, E12.5 SCG (top row) or vagal (bottom row) explants were grown in collagen gels in the presence of EDN3 (5 nM) and the EdnrA antagonist BQ123 (1 μM) or the EdnrB antagonist BQ788 (100 nM) as indicated. Quantification of EDN3-induced neurite outgrowth influenced by each antagonist is shown in Supplementary Fig. 3c, d. Scale bars, 100 μm.

in vitro results, EdnrB is not required for directed SCG axonal extension. *Ece2* mutant embryos[10] also exhibited normal SCG axonal outgrowth (data not shown). In contrast, *Ednra* mutant embryos[15] exhibited a complete lack of SCG axonal extension along the external carotid arteries (Fig. 4c, i). This axonal projection deficit is specific for the external carotid arteries; *Ednra* mutants exhibited normal SCG axonal projections along the internal carotid arteries (Fig. 4a, b, g, h). The axonal projection defect in *Ednra* mutants is not due to defective smooth muscle cell differentiation, as shown by SMA staining (Fig. 4d–f, j–l). Edn3 serves as a ligand that promotes EdnrA signalling in this context because *Edn3* mutant embryos[16] showed a similar although slightly less severe phenotype compared to that of the *Ednra* mutants, in which SCG axons failed to fully extend along the external carotid arteries (Fig. 4o, r). In *Edn3* mutants, SCG neurons did extend some processes around the carotid bifurcation, albeit to a much lesser extent than littermate controls, and they initiated a small and poorly developed projection along the proximal section of the external carotid arteries. Even at E15.5, the latest time point analysed, axonal projections from the SCG along the external carotid arteries and their branches were markedly deficient in *Edn3*$^{-/-}$ embryos (Supplementary Fig. 5). We note that *Edn1* is also expressed in the vicinity of the developing SCG, and thus Edn1 may have some functional overlap with Edn3 to support modest SCG extension along the external carotid arteries. Moreover, as observed for the *Ednra* mutants, *Edn3*-null axons extended along the internal carotid arteries and their branches normally (Fig. 4m, n, p, q and Supplementary Fig. 5).

These *in vivo* and *in vitro* findings collectively implicate Edn3 as a vascular-derived sympathetic axon growth factor. We suggest that Edn3, released from the neural-crest-derived smooth muscle layer of the external carotid arteries, directs those SCG neurons destined to innervate the salivary glands to send axonal projections along the nascent external carotid arteries—the sole pathway to these end organs. Interestingly, our analysis demonstrates that Edn3–EdnrA signalling is not required for projections to the internal carotid arteries. Because embryonic SCGs do not spontaneously exhibit axonal outgrowth, at least *in vitro*, we predict that other factors expressed by the mesoderm-derived smooth muscle layer of the internal carotid artery direct projections from the SCG towards this blood vessel. It is noteworthy that the smooth muscle layer of the external carotid arteries originates from the third pharyngeal arch neural crest. Moreover, endothelin signalling also contributes to development of additional third pharyngeal arch structures and other aspects of cardiovascular and craniofacial development[15], all of which are also dependent on neural crest cell progeny. Furthermore, endothelin receptors are broadly and dynamically expressed in these and other neural crest derivatives. Thus, it will be of interest to ask whether endothelins serve as axonal growth and guidance cues for other neural projections associated with third pharyngeal arch structures and other neural crest derivatives, such as the thymus, thyroid, cardiac outflow tracts and various craniofacial structures. Likewise, endothelins are expressed in the brain and in a variety of mesoderm-derived tissues[17] and, therefore, may serve as guidance cues for many distinct populations of developing neurons.

Our observations indicate that *Ednra* expression within the SCG is variegated, suggesting that only a subset of SCG neurons can respond to endothelins and project to the external carotid arteries. The basis of this endothelin receptor heterogeneity within the SCG is unknown; it may arise by stochastic events, by a form of lateral inhibition within the ganglion, or by predetermination within the neural crest sublineage that gives rise to SCG neurons. We do not yet know if EdnrA-positive neurons are also able to project to the internal carotid arteries although, clearly, EdnrA-negative neurons are unable to project to the external carotid arteries. Therefore, we suggest a model in which developing sympathetic neurons are predetermined to choose between distinct vascular trajectories, and are

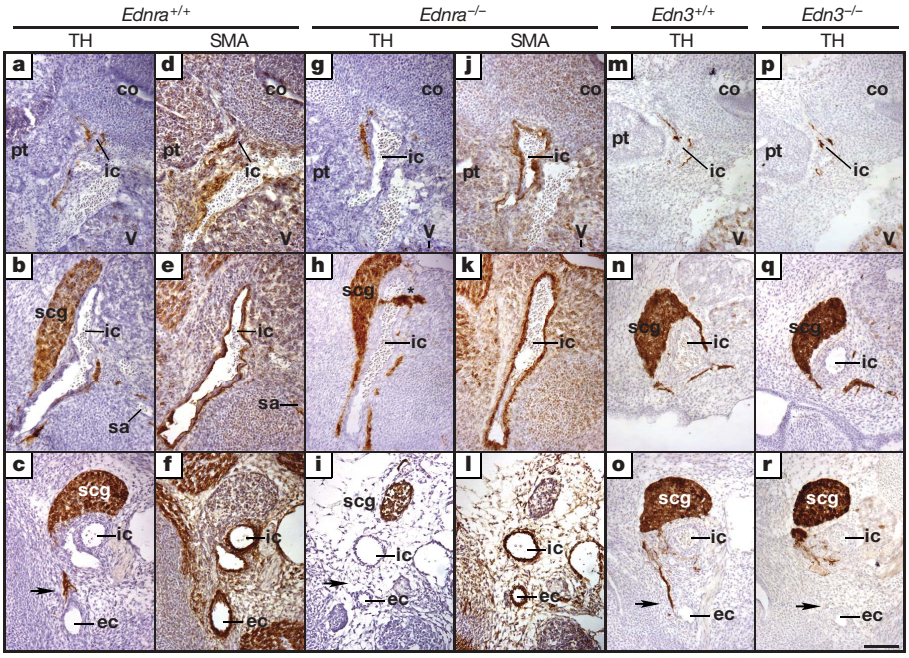


Figure 4 | Edn3 and its receptor EdnrA are essential for SCG axonal projections along the external but not internal carotid arteries in mouse embryos. **a**–**l**, Histological sections stained with TH or SMA from an E13.5 *Ednra*$^{-/-}$ embryo and a littermate control. Three mutant embryos analysed from multiple litters exhibited a complete absence of axons along the external carotid artery, similar to the embryo shown in **g**–**l**. **m**–**r**, Histological sections from an E13.5 *Edn3*$^{-/-}$ embryo and a littermate control stained with TH. Seven *Edn3*$^{-/-}$ embryos analysed at E13.5 or older from multiple litters had defects in axonal projections towards the external carotid arteries. Sections shown in the top row are all at the cochlea–pituitary level, and show the rostral projection from the SCG along the internal carotid artery. Shown in the middle row are sections at the level of the branch of the stapedial artery from the internal carotid artery; these reveal the cranial-most extent of the SCG and the initial projection from the SCG to the internal carotid artery. Sections in the lower row were taken at a level immediately above the bifurcation of the external and internal carotid arteries from the common carotid artery. Arrows point to the presence (in panels **c** and **o**) or the absence (in **i** and **r**) of axonal projections to the external carotid artery. Abbreviations: co, cochlea; pt, pituitary gland; V, trigeminal ganglion; *, histological artefact. Scale bar, 100 μm.

thus destined to innervate select end organs. This choice of vascular route is guided, at least in part, by the actions of vascular-derived endothelins.

METHODS SUMMARY

The *Wnt1cre*[18], *R26R*[19], *Ednra*[15], *spotting lethal*[14], *Edn3* (ref. 16) and *Ece2* (ref. 10) alleles have been described previously. X-gal staining and SMA immunostaining of frozen sections were as described previously[6,7]. Tyrosine hydroxylase (TH; Chemicon) immunoreactivity was visualized using an horseradish peroxidase (HRP)-conjugated secondary antibody and a diaminobenzidine (DAB) detection system. For *in situ* hybridization reactions, digoxygenin (DIG)-labelled antisense probes were generated for each gene, and the hybridization signal was detected using an alkaline phosphatase (AP)-conjugated anti-DIG antibody and 5-bromo-4-chloro-indolylphosphate/nitroblue tetrazolium (BCIP/NTB; Roche). For explant cultures, SCGs were cut into pieces on dissection, and placed into a collagen gel[11]. Explants were grown in the absence or presence of endothelins (Alexis), endothelin- or BSA-soaked heparin–agarose beads (Sigma), or the endothelin receptor antagonists BQ123 or BQ788 (AG Scientific) in Leibovitz-L15(L15)-CO_2 medium supplemented with 5% fetal bovine serum (FBS) for three days.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 November 2007; accepted 8 February 2008.

1. Inoue, A. *et al.* The human endothelin family: three structurally and pharmacologically distinct isopeptides predicted by three separate genes. *Proc. Natl Acad. Sci. USA* **86,** 2863–2867 (1989).
2. Arai, H., Hori, S., Aramori, I., Ohkubo, H. & Nakanishi, S. Cloning and expression of a cDNA encoding an endothelin receptor. *Nature* **348,** 730–732 (1990).
3. Sakurai, T. *et al.* Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. *Nature* **348,** 732–735 (1990).
4. Schotzinger, R. J. & Landis, S. C. Cholinergic phenotype developed by noradrenergic sympathetic neurons after innervation of a novel cholinergic target *in vivo*. *Nature* **335,** 637–639 (1988).
5. Hiruma, T., Nakajima, Y. & Nakamura, H. Development of pharyngeal arch arteries in early mouse embryo. *J. Anat.* **201,** 15–29 (2002).
6. Jiang, X., Rowitch, D. H., Soriano, P., McMahon, A. P. & Sucov, H. M. Fate of the mammalian cardiac neural crest. *Development* **127,** 1607–1616 (2000).
7. Choudhary, B. *et al.* Cardiovascular malformations with normal smooth muscle differentiation in neural crest-specific type II TGFβ receptor (Tgfbr2) mutant mice. *Dev. Biol.* **289,** 420–429 (2006).
8. Xu, D. *et al.* ECE-1: a membrane-bound metalloprotease that catalyzes the proteolytic activation of big endothelin-1. *Cell* **78,** 473–485 (1994).
9. Yanagisawa, H. *et al.* Dual genetic pathways of endothelin-mediated intercellular signaling revealed by targeted disruption of endothelin converting enzyme-1 gene. *Development* **125,** 825–836 (1998).
10. Yanagisawa, H. *et al.* Disruption of ECE-1 and ECE-2 reveals a role for endothelin-converting enzyme-2 in murine cardiac development. *J. Clin. Invest.* **105,** 1373–1382 (2000).
11. Guthrie, S. & Lumsden, A. Collagen gel coculture of neural tissue. *Neuroprotocols* **4,** 116–120 (1994).
12. Ihara, M. *et al. In vitro* biological profile of a highly potent novel endothelin (ET) antagonist BQ-123 selective for the ETA receptor. *J. Cardiovasc. Pharmacol.* **20** (suppl. 12), S11–S14 (1992).
13. Ishikawa, K. *et al.* Biochemical and pharmacological profile of a potent and selective endothelin B-receptor antagonist, BQ-788. *Proc. Natl Acad. Sci. USA* **91,** 4892–4896 (1994).
14. Gariepy, C. E., Cass, D. T. & Yanagisawa, M. Null mutation of endothelin receptor type B gene in spotting lethal rats causes aganglionic megacolon and white coat color. *Proc. Natl Acad. Sci. USA* **93,** 867–872 (1996).
15. Clouthier, D. E. *et al.* Cranial and cardiac neural crest defects in endothelin-A receptor-deficient mice. *Development* **125,** 813–824 (1998).
16. Baynash, A. G. *et al.* Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. *Cell* **79,** 1277–1285 (1994).
17. Firth, J. D. & Ratcliffe, P. J. Organ distribution of the three rat endothelin messenger RNAs and the effects of ischemia on renal gene expression. *J. Clin. Invest.* **90,** 1023–1031 (1992).
18. Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. & McMahon, A. P. Modification of gene activity in mouse embryos *in utero* by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8,** 1323–1326 (1998).
19. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21,** 70–71 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Deppmann, A. Kolodkin, J. Merte, S. Sockanathan and L. Schramm for comments on this manuscript, and S. Dixon and N. Murakami for technical support. This research was supported by grants from the National Institutes of Health (to H.M.S., C.E.G., M.Y. and D.D.G.). M.Y. and D.D.G. are investigators of the Howard Hughes Medical Institute.

Author Information Microarray data referred to in this study have been deposited in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE10360. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.D.G. (dginty@jhmi.edu).

METHODS

Histology and immunohistochemistry. Glutaraldehyde-fixed embryos were cryopreserved in sucrose. Five- or ten-micrometre cryosections were post-fixed, stained with X-gal by standard procedures, and counterstained with nuclear fast red. Glutaraldehyde- or paraformaldehyde-fixed embryos were cryopreserved and sectioned at 5- or 10-μm thickness for immunostaining. The primary antibodies used were TH (1:400; Chemicon) or alpha-smooth muscle actin (1:1,000; Sigma), and visualized using an HRP-conjugated secondary antibody and DAB/H_2O_2 detection system. For histological examination of sympathetic axonal phenotypes in mutant embryos, all cervical–thoracic sections were immunostained with TH. Alternatively, alternate sections from cervical to thoracic levels were subjected to hematoxylin–eosin staining and TH and SMA immunohistochemistry.

Microarray analysis. Vascular segments (the internal and external carotid arteries) were visualized by India ink injection into the left ventricle and dissected from 22 E13.5 mouse embryos, pooled, and total RNA extracted. Complementary RNAs, synthesized from complementary DNAs made from two rounds of amplification of each RNA sample, was used to hybridize and screen Affymetrix microarrays (MOE430_2) for differentially expressed genes by the JHMI Microarray Core Facility.

***In situ* hybridization.** Histological sections from glutaraldehyde-fixed Wnt1-Cre/R26R embryos were used for *in situ* hybridization reactions. Digoxygenin-labelled sense and antisense probes were generated for *Ece1* (corresponding to nucleotides 2308–2532 of NM_199307), *Edn1* (834–988 of NM_010104), *Edn2* (747–851 of NM_007902), *Edn3* (1492–1713 of NM_007903), *Ednra* (565–991 of NM_010332) and *Ednrb* (1250–1563 of NM_007904), and the hybridization signal was detected using an alkaline-phosphatase-conjugated anti-digoxygenin antibody and BCIP/NTB (Roche).

Collagen gel culture. Vascular structures in E12.5 and E13.5 embryos were visualized by injection of India ink, and SCGs and CN Xs were dissected in PBS. SCGs were cut into several pieces and placed into a collagen gel as described previously[11]. Explants were grown in the absence or presence of synthetic endothelins (EDN1, EDN2 or EDN3, Alexis), the endothelin receptor selective antagonists BQ123 or BQ788 (AG Scientific), or endothelin- or BSA-soaked heparin–agarose beads (Sigma) in L15CO_2 medium supplemented with 5% FBS for three days. Synthetic endothelins and receptor antagonists were applied at concentrations ranging from 1 nM to 100 nM and from 10 nM to 10 μM, respectively, to optimize assay conditions before performing the experiments. Neurite outgrowth was evaluated by immunostaining explants with anti-neurofilament (2H3) (1:150, Developmental Studies Hybridoma Bank) followed by immunofluorescence detection using fluorescein-conjugated secondary antibodies. To quantify neurite outgrowth, the length of neurites extending from explants was measured using an Image J image processing program.

Quantitative real-time PCR analysis. Vascular segments (the internal and external carotid arteries) were visualized by injection of India ink into the left ventricle and dissected from E12.5 mouse embryos. Total RNA was extracted from individual vascular segment by guanidinium isothiocyanate extraction and were reverse transcribed using standard conditions. All samples were analysed independently by real-time PCR using an iCycler iQ with SYBR green supermix (Bio-Rad).

LETTERS

The earliest thymic progenitors for T cells possess myeloid lineage potential

J. Jeremiah Bell[1] & Avinash Bhandoola[1]

There exists controversy over the nature of haematopoietic progenitors of T cells. Most T cells develop in the thymus, but the lineage potential of thymus-colonizing progenitors is unknown. One approach to resolving this question is to determine the lineage potentials of the earliest thymic progenitors (ETPs). Previous work has shown that ETPs possess T and natural killer lymphoid potentials, and rare subsets of ETPs also possess B lymphoid potential[1], suggesting an origin from lymphoid-restricted progenitor cells. However, whether ETPs also possess myeloid potential is unknown. Here we show that nearly all ETPs in adult mice possess both T and myeloid potential in clonal assays. The existence of progenitors possessing T and myeloid potential within the thymus is incompatible with the current dominant model of haematopoiesis, in which T cells are proposed to arise from lymphoid-restricted progenitors[2]. Our results indicate that alternative models for lineage commitment during haematopoiesis must be considered.

Haematopoiesis is a model system for tissue homeostasis and regeneration. Mature blood cell types originate from haematopoietic stem cells (HSCs) that are multipotent and self-renewing[3]. The first step in differentiation of HSCs involves the generation of multipotent progenitors (MPPs), which do not self-renew[4,5]. This is followed by an ordered series of irreversible lineage commitment events, resulting in the generation of increasingly lineage-restricted downstream intermediates. Subsequent to the generation of non-renewing MPPs, the standard classical model of haematopoiesis postulates a binary split between lymphoid and myelo-erythroid lineages[2,6,7]. In this model, MPPs give rise to either common lymphoid progenitors (CLPs) with lymphoid potential but devoid of myeloid and erythroid potentials, or to common myeloid progenitors possessing myeloid and erythroid potentials but lacking lymphoid potentials[6,7]. Further development along a given lineage is accompanied by continuing loss of alternative lineage potentials, until mature blood cell types are generated[2].

As lymphocytes, T cells are proposed to develop from CLPs[7]. However, the identity of thymus-settling progenitors is unknown[1]. The most immature intrathymic progenitors that have been identified are ETPs, defined phenotypically as expressing low levels of lineage markers (Lin) that are also $CD25^{-}Kit^{hi}$ cells[8,9]. In addition to efficient T lymphoid potential[9,10], ETPs possess natural killer and dendritic cell potentials[1]. These non-T-lineage potentials are maintained in downstream DN2 cells ($Lin^{-}CD25^{+}Kit^{hi}$), and are not lost until final commitment to the T-cell lineage, which occurs at the DN3 stage ($Lin^{-}CD25^{+}Kit^{-/lo}$)[1]. In addition to T, natural killer and dendritic cell potential, a rare subset of ETPs also possesses B lymphoid potential[11–13]. Whether ETPs possess potential for myeloid cells such as monocytes/macrophages and granulocytes, as distinct from their well-studied dendritic cell potential, is less clear. The ETP population possesses some degree of myeloid potential in bulk assays[9,14–16], leading to the suggestion that not all ETPs are derived from lymphoid-restricted CLPs[9]. It is unclear whether the same cells that have myeloid potential also have T lymphoid potential, or if the ETP population instead contains some myeloid-lineage-committed cells (Supplementary Fig. 1)[14]. Resolution of this question requires the development of clonal assays to measure the myeloid and T lineage potentials of single ETPs.

Culture of haematopoietic progenitors on bone-marrow-derived OP9 stromal cells has been extensively used to investigate lineage potentials of haematopoietic progenitors. OP9 stromal cells transduced with the Notch ligand delta-like 4 (OP9-DL4) support T-lineage development, whereas OP9 cells lacking delta-like Notch ligands (OP9) do not[17]. Using this system we found that ETPs placed in short-term cultures with Flt3 ligand and interleukin (IL)-7 underwent population expansion (Fig. 1a). Similar levels of expansion were seen when ETPs were cultured on OP9-DL4 or on OP9 stroma.

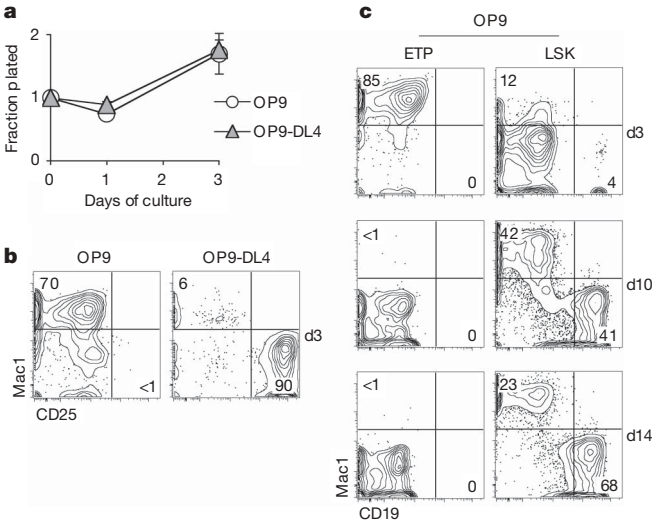


Figure 1 | ETPs have myeloid lineage potential. **a**, Sorted ETPs (500 per well) were cultured with either OP9 or OP9-DL4 stromal cells in medium supplemented with IL-7 and Flt3 ligand. Cultures were analysed at 1 and 3 days and assessed for the presence of live (4,6-diamidino-2-phenylindole (DAPI) negative) haematopoietic ($CD45^{+}$) cells. Recovery values are shown as the fraction of cells plated ± s.e.m. of triplicate wells. **b**, Cells cultured for 3 days were analysed for surface expression of Mac1 and CD25 by flow cytometry. Plots shown are gated on live haematopoietic cells. Numbers on plots represent percentages of cells in those regions. Most Mac1-negative cells lacked expression of lineage markers and so were undifferentiated cells. **c**, Mac1 and CD19 expression on cells derived from ETPs or bone marrow LSK cells (500 per well) cultured for 3, 10, or 14 days on OP9 stroma. Data in **a**–**c** are representative of at least three experiments.

[1]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

However, the identities of the resulting cell lineages were distinct. On OP9-DL4 stroma, ETPs gave rise to CD25$^+$ T-lineage cells (Fig. 1b), which co-expressed Thy-1 and other T-lineage-specific markers[17]. By day 3 on OP9 stroma, however, ETPs gave rise to cells that expressed the myeloid lineage marker Mac1 (Fig. 1b). Such Mac1$^+$ cells were also seen when DN2 cells were cultured, albeit in lower numbers, but not in cultures initiated with DN3 cells (Supplementary Fig. 2). B-lineage cells were not seen in these experiments initiated with 500 ETPs on OP9 stroma (Fig. 1c), even when cultures were extended until 2 weeks. This can be explained by the low frequency of ETPs with B potential (<1:500 in 6-week-old adult C57BL/6 mice)[12,13,18]. In contrast, primitive bone marrow Lin$^-$Sca1$^+$Kit^{hi} (LSK) progenitors efficiently generated B-lineage cells in these culture conditions, as evidenced by the presence of CD19$^+$ cells from day 3 until 2 weeks of culture (Fig. 1c).

We further characterized the Mac1$^+$ progeny of ETPs cultured on OP9 stroma (Fig. 2). Some Mac1$^+$CD11c$^+$ dendritic cells were evident in these cultures, as expected[11,19] (Fig. 2a). However, the majority of Mac1$^+$ cells lacked expression of CD11c. Some Mac1$^+$ cells co-expressed the monocyte/macrophage markers MCSF receptor (MCSFR) and F4/80. Other Mac1$^+$ cells co-expressed Gr1, expressed on granulocytes and some macrophages. To characterize better these populations, we made use of lysozyme M reporter mice, in which the gene for green fluorescent protein (GFP) is knocked into the

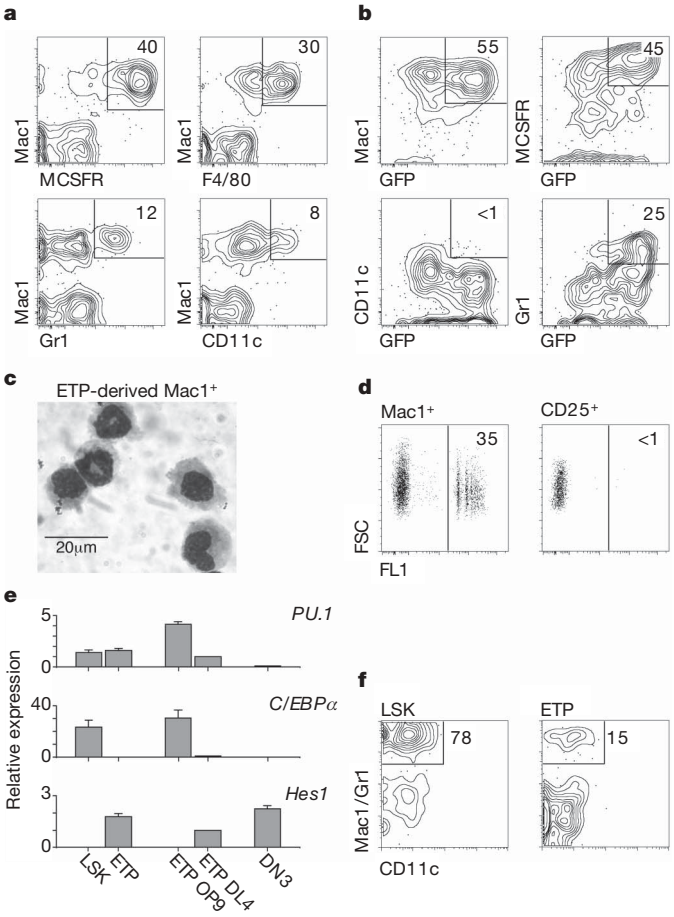


Figure 2 | Characterization of ETP-derived myeloid cells. **a**, ETPs were cultured for 3 days on OP9 stroma in medium supplemented with IL-7 and Flt3 ligand, and their progeny were assessed for myeloid lineage markers by flow cytometry. **b**, ETPs were sorted from lysozyme M reporter mice (where GFP is knocked into the lysozyme M locus) and cultured with OP9 stromal cells as before. After 3 days, cells were analysed for expression of myeloid markers and GFP. **c**, Mac1$^+$ cells resulting from ETPs cultured as in **a** were Wright-Giemsa stained. Scale bar, 20 μm. Data in **a**–**c** are representative of at least three experiments. **d**, Phagocytic ability of sorted Mac1$^+$ cells derived from ETPs cultured on OP9 was analysed by uptake of beads fluorescing in the FL1 channel (see Methods). CD25$^+$ cells from ETPs cultured on OP9-DL4 serve as a negative control. FSC, forward scatter. Data are representative of two experiments. **e**, ETP-derived cultures on OP9 stroma (ETP OP9) or OP9-DL4 stroma (ETP DL4) were harvested after 4 days, then analysed by real-time PCR for expression of *PU.1*, *C/EBPα* and *Hes1*. Freshly isolated bone marrow LSK progenitors, ETPs and DN3 cells served as controls. Data are averages of triplicate wells ± s.e.m., and are representative of two experiments. **f**, ETPs or control bone marrow LSK progenitors (1,000) from C57BL/6 mice (CD45.2) were transferred intravenously into irradiated B6.CD45.1 recipients. After 6 days the spleens were analysed for donor-derived cells using CD45 markers as shown in Supplementary Fig. 3. Mac1/Gr1 versus CD11c expression on gated donor-derived cells is shown. Data are representative of five experiments using 200–3,000 progenitors in which myeloid lineage progeny were always present.

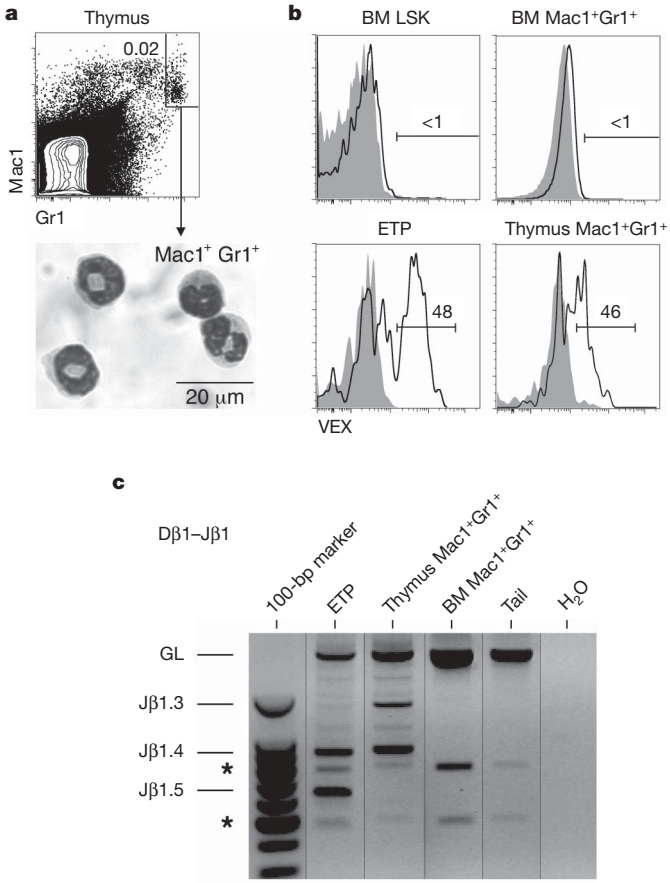


Figure 3 | Thymic granulocytes possess antigen receptor gene rearrangements. **a**, Thymic Mac1$^+$Gr1$^+$ cells were isolated by cell sorting and Wright-Giemsa stained. Band, ring and segmented granulocytes were present. Scale bar, 20 μm. Data are representative of two experiments. **b**, Bone marrow (BM) and thymic populations from H2-SVEX fluorescent transgenic recombination reporter mice were analysed for VEX fluorescence by flow cytometry. Plots are shown comparing H2-SVEX bone marrow and thymic progenitor and granulocyte populations (open histograms) with corresponding C57BL/6 control populations (shaded histograms). The gating strategies for analysis of these populations are shown in Supplementary Fig. 4. Data are representative of four experiments. **c**, Wild-type bone marrow and thymic granulocyte populations were analysed for Dβ1–Jβ1 rearrangements by PCR. ETPs, previously shown to possess such rearrangements, are also shown. The identities of marked Dβ1–Jβ1 rearrangements were confirmed by sequencing (Supplementary Fig. 5). The germline band (GL) is indicated. Background bands marked with an asterisk are present in all samples, including tail. DNA sequencing was used to confirm that these bands do not contain rearranged TCR or immunoglobulin genes (data not shown). No consistent difference was seen in TCR rearrangements within ETPs and thymic granulocytes. Data are representative of three experiments.

lysozyme M locus[20,21]. HSCs and ETPs express only low levels of GFP or are GFP-negative[21] (our own unpublished data). High-level GFP expression is seen in monocytes/macrophages and granulocytes, but not in dendritic cells or other non-myeloid lineages[20] (Fig. 2b). $Mac1^+$ cells developing from ETPs included GFP^{hi} cells. Indeed, the $MCSFR^+$ cells and $Gr1^+$ cells were mainly GFP^{hi}, whereas $CD11c^+$ dendritic cells lacked high-level expression of GFP (Fig. 2b). Microscopic examination of $Mac1^+$ cells derived from ETPs revealed some cells with monocytic morphology and other cells with granulocytic morphology (Fig. 2c). Some $Mac1^+$ cells phagocytosed fluorescent latex beads, indicating that they were functional macrophages (Fig. 2d). These results further establish that $Mac1^+$ progeny of ETPs include macrophages and granulocytes that are in the myeloid lineage.

We also determined that these cells possessed transcripts consistent with their myeloid lineage identity. Bone marrow LSK progenitors known to possess myeloid potential[5] and ETPs were compared for expression of *PU.1* (also called *Sfpi1*), *C/EBPα* (*Cebpa*) and the Notch gene target *Hes1*. Withdrawal of Notch signalling caused increased *PU.1* and *C/EBPα* expression in progeny of ETPs, consistent with the myeloid identity of cells in these cultures[22]. Maintenance of Notch signalling prevented these transcriptional changes (Fig. 2e). To determine whether the myeloid potential of ETPs was evident *in vivo*, we intravenously transferred 1,000 ETPs or bone marrow LSK cells into irradiated mice, and analysed spleens for donor-derived $Mac1/Gr1^+$ populations 6 days later. The overall level of donor chimaerism was similar between mice inoculated with ETPs and LSK cells at this early time point (Supplementary Fig. 3). Although the frequency of myeloid progeny was lower for mice receiving ETP donor cells compared with LSK donor cells, ETPs gave rise to myeloid cells in this experimental *in vivo* system (Fig. 2f).

We next wished to determine whether ETPs physiologically give rise to intrathymic myeloid progeny *in vivo*. We focused on $Mac1^+Gr1^+$ cells because we could easily distinguish this population even at the low frequency present within the thymus (Fig. 3a). Thymic $Mac1^+Gr1^+$ cells possessed a granulocytic morphology and were unambiguously myeloid lineage cells (Fig. 3a). To mark non-lymphoid progeny of ETPs permanently, we made use of a fluorescent transgenic V(D)J recombination reporter mouse[23]. This transgene is expressed in haematopoietic cells in the H2-SVEX mouse[23]. Rearrangement mediated by recombinase-activating genes (RAG) results in permanent expression of violet excited GFP (VEX), allowing detection by flow cytometry[23]. We found that bone marrow LSK progenitors and bone marrow $Mac1^+Gr1^+$ cells do not express VEX (Fig. 3b). In contrast, nearly half of ETPs express VEX, consistent with their possession of T-cell receptor (TCR) Dβ1–Jβ1 as well as additional rearrangements[9]. A fraction of thymic granulocytes also exhibited VEX fluorescence, demonstrating that they were derived from RAG-expressing progenitors. We confirmed that thymic granulocytes possessed Dβ1–Jβ1 rearrangements, similar to ETPs (Fig. 3c). The identity of Dβ1–Jβ1 rearrangements in thymic granulocytes was confirmed by DNA sequencing (Supplementary Fig. 5). A smaller fraction (5–10%) of the $Mac1^+MCSFR^+CD11c^-$ thymic macrophage population was also positive for VEX in H2-SVEX mice (data not shown), suggesting that other myeloid cell types might also be derived from ETPs. These results indicate that ETPs physiologically give rise to myeloid cells *in vivo*.

The prominent myeloid potential of ETPs is surprising, and suggests that this potential is not due to a rare myeloid committed progenitor. Instead, bipotent progenitors possessing T and myeloid potential may exist within the ETP population. To test this, we performed single cell assays. As ETPs developed predominantly into T-lineage cells on OP9-DL4 stroma, and developed into myeloid lineage cells on OP9 stroma, we cultured ETPs on a 50:50 mixture of OP9:OP9-DL4 for 10 days. Our rationale was that as lineage commitment is not instantaneous[11,24], uncommitted progeny of ETPs would segregate to different local environments in these cultures, and adopt different cell fates. The inclusion of cytokines promoting myeloid lineage and T-lineage development allowed expansion of myeloid lineage and T-lineage progeny of single ETPs to a degree that could be detected by flow cytometry. Unlike 3-day cultures where granulocytic progeny were abundant (Fig. 2c), most myeloid cells in these 10-day cultures were monocytes/macrophages (data not shown). This is consistent with the accompanying study[25] that examined longer-term cultures, in which macrophages also predominated. The *in vitro* system established here allowed analysis of the myeloid lineage and T-lineage potentials of single ETPs. Surprisingly, most ETPs contained both T and myeloid potential (Table 1 and Supplementary Fig. 6). Some progenitors with T and myeloid potential remained evident at the downstream DN2 stage; however, the majority of DN2 cells lacked myeloid potential (Table 1). Hence, myeloid potential is evident in most ETPs, is largely but not completely lost in DN2 cells, and is extinguished in DN3 cells (Supplementary Fig. 2).

Our results establish that most ETPs possess T and myeloid potential, and so establish that the majority of ETPs do not arise from lymphoid-restricted progenitors. Instead, we suggest that ETPs arise from multipotent bone marrow progenitors such as early lymphoid progenitors or lymphoid-primed multipotent progenitors, which possess T, B and myeloid potential[18,26–28]. We propose that such multipotent progenitor cells lose B potential, either prethymically or intrathymically, generating ETPs possessing T and myeloid potential (Supplementary Fig. 1). Our results further establish that whereas myeloid potential is lost in the bone marrow during B-cell development, it is lost in the thymus during T-cell development, suggesting that distinct mechanisms may be involved. The myeloid potential of ETPs was most clearly evident when ETPs were removed from the thymus and assessed in the absence of T-inductive Notch signalling, indicating a role for intrathymic Notch signals in suppression of the myeloid fate. However, we also found that granulocytes within the thymus can possess TCR rearrangements, suggesting that some ETPs produce myeloid progeny within the thymus. The development of such thymic myeloid cells may have commonalities with the previously described development of thymic dendritic cells from ETPs[19]. Notch signals antagonize the effects of PU.1 and C/EBPα, critical for development of dendritic cells and other myeloid lineages, suggesting that some thymic dendritic cells arise from ETPs in niches protected from Notch signalling[22,29]. Similarly, thymic myeloid cells may arise from ETPs that escape Notch signals. Thus, the development of dendritic cells and myeloid cells within the thymus may share some cellular requirements and molecular mechanisms.

The standard model of a binary split between lymphoid and myelo-erythroid lineages has been the prevailing idea of initial lineage restriction during haematopoiesis[6,7]. The present study, together with past work indicating that the majority of ETPs lack B potential[9,11–13,16,30], demonstrates that most ETPs possess T-lymphoid and myeloid potentials but lack B-lymphoid potential. The standard model cannot explain the lineage potentials of ETPs, because this model requires that myeloid potential be lost before the loss of any lymphoid potentials during T-cell development. Hence, the idea of

Table 1 | A high frequency of single ETPs have T and myeloid potential

Cell type	T (%)	T and myeloid (%)	Myeloid (%)
ETP	9	87	4
DN2	87	13	0

Single ETPs or DN2 cells from lysozyme M reporter mice were sorted into 96-well plates containing a 50:50 mix of OP9 and OP9-DL4, and medium supplemented with 1 ng ml^{-1} IL-7, 5 ng ml^{-1} Flt3 ligand and 10 ng ml^{-1} IL-3, IL-6, stem cell factor (SCF), macrophage colony-stimulating factor (M-CSF), granulocyte colony-stimulating factor (G-CSF), and granulocyte/macrophage colony-stimulating factor (GM-CSF). After 10 days the resulting colonies were analysed for expression of T (CD25) and myeloid (Mac1, MCSFR, Gr1 and GFP) markers as well as CD45 and CD11c. Shown is the percentage of T or myeloid unipotent and T and myeloid bipotent progenitors. The percentages are normalized based on the number of surviving colonies arising from single cells, excluding wells that did not support growth. Plating efficiencies were >70%. Colonies derived from 150 single ETPs (four experiments) and 55 single DN2 cells (two experiments) were analysed. All experiments gave similar results.

an obligatory initial split in haematopoiesis between lymphoid lineages and myelo-erythroid lineages[6,7] is incorrect.

METHODS SUMMARY

C57BL/6 (B6) mice and B6.CD45.1 mice were purchased from the National Cancer Institute. Lysozyme M reporter mice[21] were provided by T. Graf. H2-SVEX mice[23] were obtained from R. M. Gerstein. Progenitor populations were isolated as described[9]. OP9 and OP9-DL4 cells were gifts from J. C. Zuniga-Pflucker and were used as described[17]. IL-7 was added at a final concentration of 1 ng ml^{-1} and Flt3 ligand at 5 ng ml^{-1}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 October 2007; accepted 15 February 2008.

1. Bhandoola, A., von Boehmer, H., Petrie, H. T. & Zuniga-Pflucker, J. C. Commitment and developmental potential of extrathymic and intrathymic T cell precursors: plenty to choose from. *Immunity* **26,** 678–689 (2007).
2. Kondo, M., Scherer, D. C., King, A. G., Manz, M. G. & Weissman, I. L. Lymphocyte development from hematopoietic stem cells. *Curr. Opin. Genet. Dev.* **11,** 520–526 (2001).
3. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241,** 58–62 (1988).
4. Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E. & Weissman, I. L. Identification of a lineage of multipotent hematopoietic progenitors. *Development* **124,** 1929–1939 (1997).
5. Adolfsson, J. *et al.* Upregulation of Flt3 expression within the bone marrow Lin$^-$Sca1$^+$c-kit$^+$ stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* **15,** 659–669 (2001).
6. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404,** 193–197 (2000).
7. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91,** 661–672 (1997).
8. Shortman, K. & Wu, L. Early T lymphocyte progenitors. *Annu. Rev. Immunol.* **14,** 29–47 (1996).
9. Allman, D. *et al.* Thymopoiesis independent of common lymphoid progenitors. *Nature Immunol.* **4,** 168–174 (2003).
10. Min, H., Montecino-Rodriguez, E. & Dorshkind, K. Reduction in the developmental potential of intrathymic T cell progenitors with age. *J. Immunol.* **173,** 245–250 (2004).
11. Benz, C. & Bleul, C. C. A multipotent precursor in the thymus maps to the branching point of the T versus B lineage decision. *J. Exp. Med.* **202,** 21–31 (2005).
12. Sambandam, A. *et al.* Notch signaling controls the generation and differentiation of early T lineage progenitors. *Nature Immunol.* **6,** 663–670 (2005).
13. Ceredig, R., Bosco, N. & Rolink, A. G. The B lineage potential of thymus settling progenitors is critically dependent on mouse age. *Eur. J. Immunol.* **37,** 830–837 (2007).
14. Wu, L., Antica, M., Johnson, G. R., Scollay, R. & Shortman, K. Developmental potential of the earliest precursor cells from the adult mouse thymus. *J. Exp. Med.* **174,** 1617–1627 (1991).
15. Matsuzaki, Y. *et al.* Characterization of c-kit positive intrathymic stem cells that are restricted to lymphoid differentiation. *J. Exp. Med.* **178,** 1283–1292 (1993).
16. Balciunaite, G., Ceredig, R. & Rolink, A. G. The earliest subpopulation of mouse thymocytes contains potent T, significant macrophage, and natural killer cell but no B-lymphocyte potential. *Blood* **105,** 1930–1936 (2005).
17. Schmitt, T. M. & Zuniga-Pflucker, J. C. T-cell development, doing it in a dish. *Immunol. Rev.* **209,** 95–102 (2006).
18. Lu, M. *et al.* The earliest thymic progenitors in adults are restricted to T, NK, and dendritic cell lineage and have a potential to form more diverse TCRβ chains than fetal progenitors. *J. Immunol.* **175,** 5848–5856 (2005).
19. Ardavin, C., Wu, L., Li, C. L. & Shortman, K. Thymic dendritic cells and T cells develop simultaneously in the thymus from a common precursor population. *Nature* **362,** 761–763 (1993).
20. Faust, N., Varas, F., Kelly, L. M., Heck, S. & Graf, T. Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* **96,** 719–726 (2000).
21. Ye, M. *et al.* Hematopoietic stem cells expressing the myeloid lysozyme gene retain long-term, multilineage repopulation potential. *Immunity* **19,** 689–699 (2003).
22. Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 transcription factors. *Immunity* **25,** 731–744 (2006).
23. Borghesi, L. *et al.* B lineage-specific regulation of V(D)J recombinase activity is established in common lymphoid progenitors. *J. Exp. Med.* **199,** 491–502 (2004).
24. Taghon, T. N., David, E. S., Zuniga-Pflucker, J. C. & Rothenberg, E. V. Delayed, asynchronous, and reversible T-lineage specification induced by Notch/Delta signaling. *Genes Dev.* **19,** 965–978 (2005).
25. Wada, H. *et al.* Adult T-cell progenitors retain myeloid potential. *Nature* doi:10.1038/nature06839 (this issue).
26. Igarashi, H., Gregory, S., Yokota, T., Sakaguchi, N. & Kincade, P. Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow. *Immunity* **17,** 117–130 (2002).
27. Adolfsson, J. *et al.* Identification of Flt3$^+$ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell* **121,** 295–306 (2005).
28. Yoshida, T., Ng, S. Y., Zuniga-Pflucker, J. C. & Georgopoulos, K. Early hematopoietic lineage restrictions directed by Ikaros. *Nature Immunol.* **7,** 382–391 (2006).
29. Franco, C. B. *et al.* Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. *Proc. Natl Acad. Sci. USA* **103,** 11993–11998 (2006).
30. Porritt, H. E. *et al.* Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages. *Immunity* **20,** 735–745 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Bassing, M. Carroll, B. Schwarz and J. Thompson for advice. This work was supported by a grant from the NIH and a Scholar Award from the Leukemia and Lymphoma Society (A.B.) and an institutional training grant from NCI (J.J.B).

Author Contributions J.J.B. performed the majority of the experimental work. J.J.B. and A.B. together planned the project, analysed data and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.B. (bhandooa@mail.med.upenn.edu).

METHODS

Mice. All mice were females used at 5–8 weeks of age. All animal experiments were done according to protocols approved by the Office of Regulatory Affairs of the University of Pennsylvania in accordance with guidelines set forth by the National Institutes of Health.

Flow cytometry. All antibodies were purchased from BD Biosciences or Ebiosciences. Antibodies used were fluorescein isothiocyanate (FITC)-conjugated CD45.1 (A20); phycoerythrin (PE)-conjugated CD25 (PC61), MCSFR (AFS98), F4/80 (BM8); PE-Cy7-conjugated Mac1 (CD11b) (M1/70); PE-Cy5.5-conjugated CD45.1, CD45.2 (104); peridinine-chlorophyll protein (PerCP)-conjugated CD19 (1D3); allophycocyanin (APC)-conjugated Kit (CD117) (2B8), CD25, CD11c (HL3); APC-Alexa-750-conjugated CD25; and biotin-conjugated CD45.2 (104), Gr1 (8C5). To exclude mature cells, a lineage cocktail of the following FITC-, PE-, or biotin-conjugated antibodies was used: B220 (RA3-6B2), CD19, CD11b, Gr1, CD11c, NK1.1 (PK136), Ter119, CD8α (53-6.7), CD8β (53-5.8), TCR-γδ (GL-3), TCR-β (H57) and CD3 (2C11). Biotin-conjugated antibodies were visualized with streptavidin-Pacific blue (Molecular Probes) or streptavidin-PerCP–Cy5.5 (BD Biosciences). Dead cells were excluded with 4,6-diamidino-2-phenylindole (DAPI). Cells were sorted on a BD FACSAria (Becton Dickinson) and sort purity was routinely checked. Cell analysis was done on a 4-laser LSR II (Becton Dickinson) and data were analysed using FlowJo version 4.6.2 (Tree Star).

Phagocytosis assay. ETP-derived $Mac1^+$ cells from OP9 cultures supplemented with M-CSF, GM-CSF, G-CSF, IL-3 and IL-6 to support maturation were sorted and incubated with 1 μm fluorescent beads (Fluoresbrite Yellow Green Microspheres, Polysciences, Inc.) for 5 h in medium supplemented with M-CSF. Control $CD25^+$ cells were obtained from OP9-DL4 cultures. Cell to bead ratio was 1:10. Cells were then treated with 0.05% trypsin-EDTA to free surface-bound beads and washed with PBS. Beads that had not been ingested were removed by centrifugation (100*g* for 10 min) on a cushion of PBS-3% BSA supplemented with 4.5% glucose. Fluorescence was then assessed by flow cytometry as above.

PCR. For real-time PCR, messenger RNA from sorted populations was isolated using the RNeasy kit (Qiagen) and reverse transcribed with Superscript II (Invitrogen). Resultant complementary DNAs were then amplified and detected using pre-made FAM-labelled primer/probes against *PU.1*, *C/EBPα* and *Hes1*, purchased from Applied Biosystems. Amplification and analysis were carried out on a 7300 Real-Time PCR System (Applied Biosystems). PCR analysis for TCR Dβ1–Jβ1 rearrangements using 1,000 cell equivalents was carried out as described[9].

Wright-Giemsa staining. Sorted populations were spun onto glass slides using a Shandon Cytospin 2 cytocentrifuge. Cells were then fixed in fresh methanol and stained in Wright-Giemsa reagent (Fisher Scientific) for 3 min, followed by Wright-Giemsa with Original Azure Blend (Harleco) for 10 min, then Wright-Giemsa with phosphate buffer pH 6.8 (Fisher Scientific) for 2 min. Stained slides were then washed with ddH_2O and allowed to dry, coverslipped, and examined under microscope. Microscopy pictures shown are at ×100 magnification.

OP9 and OP9-DL4 cell culture. OP9 and OP9-DL4 cells were used essentially as described[17]. For most cultures, IL-7 was added at a final concentration of 1 ng ml^{-1} and Flt3 ligand at 5 ng ml^{-1}. Myeloid progeny of ETPs cultured in these conditions included granulocytes and monocytes/macrophages that were evident at day 3 of culture. Myeloid progeny of ETPs did not survive to day 10 of culture in the absence of additional cytokine supplementation (shown in Fig. 1). For single cell assays, ETPs were cultured on a 50:50 mixture of OP9:OP9-DL4 for 10 days. IL-3, IL-6, stem cell factor, macrophage colony-stimulating factor, granulocyte colony-stimulating factor, and granulocyte/macrophage colony-stimulating factor were added at a final concentration of 10 ng ml^{-1}, in addition to IL-7 and Flt3 ligand as before. These myeloid-cytokine-supplemented 10-day cultures contained abundant monocytes/macrophages, but only rare granulocytes. The relative paucity of granulocytes compared to monocytes/macrophages after extended periods of *in vitro* culture is probably a result of the shorter lifespan of mature granulocytes. Stromal cells were plated 2 days before initiation of culture at a concentration of 20,000 cells ml^{-1} in 24-well plates for bulk cultures and 96-well plates for single cell assays. For bulk cultures, 500 progenitors were plated per well. For single cell assessment of ETPs, 150 colonies were analysed.

LETTERS

Adult T-cell progenitors retain myeloid potential

Haruka Wada[1]†, Kyoko Masuda[1]†, Rumi Satoh[1], Kiyokazu Kakugawa[1], Tomokatsu Ikawa[1], Yoshimoto Katsura[1,2] & Hiroshi Kawamoto[1]

During haematopoiesis, pluripotent haematopoietic stem cells are sequentially restricted to give rise to a variety of lineage-committed progenitors. The classical model of haematopoiesis postulates that, in the first step of differentiation, the stem cell generates common myelo-erythroid progenitors and common lymphoid progenitors (CLPs). However, our previous studies in fetal mice showed that myeloid potential persists even as the lineage branches segregate towards T and B cells[1–6]. We therefore proposed the 'myeloid-based' model of haematopoiesis[7,8], in which the stem cell initially generates common myelo-erythroid progenitors and common myelo-lymphoid progenitors. T-cell and B-cell progenitors subsequently arise from common myelo-lymphoid progenitors through myeloid-T and myeloid-B stages, respectively. However, it has been unclear whether this myeloid-based model is also valid for adult haematopoiesis. Here we provide clonal evidence that the early cell populations in the adult thymus contain progenitors that have lost the potential to generate B cells but retain substantial macrophage potential as well as T-cell, natural killer (NK)-cell and dendritic-cell potential. We also show that such T-cell progenitors can give rise to macrophages in the thymic environment *in vivo*. Our findings argue against the classical dichotomy model in which T cells are derived from CLPs; instead, they support the validity of the myeloid-based model for both adult and fetal haematopoiesis.

By using a clonal analysis of lympho-haematopoietic cells in fetal mice, we previously showed the existence of bipotent progenitors of myeloid-T and myeloid-B lineages, in addition to multipotent and unipotent progenitors[1–5]. However, our clonal analysis failed to detect any of the CLP-type progenitors. More recently we discovered the presence of common myelo-lymphoid progenitors in fetal liver[6], and thus came to propose a myeloid-based model of haematopoiesis[7,8], in which myeloid potential is retained in erythroid, T-cell and B-cell branches (Supplementary Fig. 1). By contrast, CLPs have been reported to be present in a bone marrow subpopulation in the adult mouse[9]. To reconcile these disparate findings between fetal liver and adult bone marrow, the concept has emerged that fetal haematopoiesis differs from adult haematopoiesis, with the CLP stage existing only during adult haematopoiesis[10–13].

We decided to examine critically whether the CLP stage exists or not in the developmental pathway from the stem cell to T cells during adult haematopoiesis. Previous studies have reported that a subpopulation of adult thymus cells that has no B-cell potential still retains myeloid and T-cell potential[14–16]. These findings suggest that B-cell potential is shut off earlier than myeloid potential, providing support for the myeloid-based model. However, lacking clonal evidence, these studies failed to provide sufficient impact to change the existing model in haematopoiesis.

We first examined whether early thymic progenitors are able to produce granulocytes and/or macrophages. The $CD4^-CD8^-$ (double-negative) fraction of adult thymus cells was subdivided into $c\text{-}Kit^+CD25^-$, $c\text{-}Kit^+CD25^+$, $c\text{-}Kit^-CD25^+$ subsets, which are designated as DN1, DN2 and DN3 subsets, respectively (Fig. 1a). We cultured DN1, DN2 and DN3 cells from adult thymus, and lineage-marker $(Lin)^-c\text{-}Kit^+Sca\text{-}1^+$ (LKS) cells from bone marrow

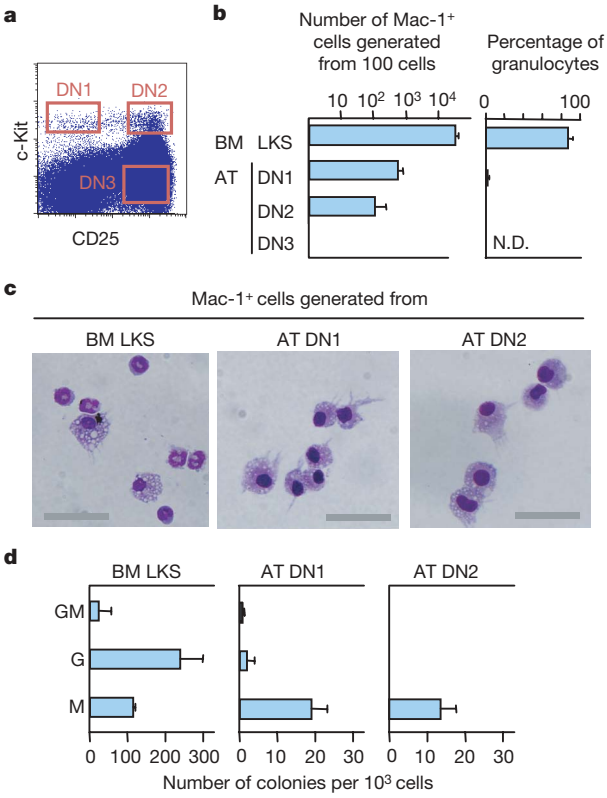


Figure 1 | Early progenitors in the adult thymus retain the potential to generate macrophages. **a**, The c-Kit/CD25 profile of the Lin^- fraction of adult thymus (AT) cells from eight-week-old mice. **b**, **c**, Cells from the indicated subpopulations of AT and LKS bone marrow (BM) cells were cultured (100 cells per well) for seven days with the PA6 stromal cell line in the presence of granulocyte colony-stimulating factor. Numbers of $Mac\text{-}1^+$ cells per well (mean and s.d. for triplicate cultures) are shown (**b**, left panel). Generated $Mac\text{-}1^+$ cells were sorted, spun onto glass slides, and Wright-stained. The proportion of cells with granulocytic morphology (mean and s.d. for triplicate culture) (**b**, right panel) and representative photomicrographs of $Mac\text{-}1^+$ cells (**c**) are shown. Scale bars, 50 μm. **d**, A CFU-C assay was performed with DN1 and DN2 AT cells and also with LKS bone marrow cells. Colony numbers (mean and s.d. for triplicate cultures) per 10^3 cells are shown. GM, colonies containing granulocytes and macrophages; G, granulocyte colonies; M, macrophage colonies.

[1]Laboratory for Lymphocyte Development, RIKEN Research Center for Allergy and Immunology, Yokohama 230-0045, Japan. [2]Division of Cell Regeneration and Transplantation, Advanced Medical Research Center, Nihon University School of Medicine, Tokyo 173-8610, Japan. †Present addresses: Division of Bioregulation Research, Institute of Medical Science, St Marianna University School of Medicine, Kawasaki 216-8512, Japan (H.W.); Institute of Molecular Medicine and Genetics, University of Georgia, GA 30602, USA (K.M.).

as a control, at 100 cells per well, on a monolayer of the stromal cell line PA6, which preferentially supports the generation of myeloid cells[17]. Mac-1$^+$ cells were generated in cultures of DN1 and DN2 cells, although many fewer cells were recovered than from cultures of LKS cells (Fig. 1b). Although LKS progenitors gave rise mainly to granulocytes, only a few were generated from DN1 cells, and no granulocytes at all were generated from DN2 cells (Fig. 1b, c). The results were confirmed by a colony-forming units in culture (CFU-C) assay (Fig. 1d). Similar findings were obtained with fetal thymus progenitors (Supplementary Fig. 2). These results indicate that the myeloid potential of thymic progenitors is primarily specified to macrophages. Expression of the transcription factor PU.1 (also called Sfpi1) has been implicated in the myeloid potential of early T-cell progenitors[18]. Indeed, RT–PCR analysis confirmed that DN1 and DN2 cells express PU.1, but PU.1 is downregulated at the DN3 stage (Supplementary Fig. 3).

To examine T-cell and myeloid potential of thymic progenitors at a single-cell level, we developed a clonal assay system with the stromal cell line TSt-4/DLL1 (ref. 19), which supports the generation of both T cells and myeloid cells (Fig. 2a). As a source of progenitor cells, we used green fluorescent protein (GFP)-transgenic mice ('green mouse')[20]. In this green mouse strain, macrophages and dendritic cells express GFP highly, whereas differentiated T cells are GFP$^-$, making the detection of GFP$^+$ macrophages easier. When 192 DN1 cells were individually cultured with TSt-4 cells, 18 wells containing only macrophages were detected and no B-cell-containing wells were observed (Fig. 2b, middle panel). In co-cultures with TSt-4/DLL1 cells, 136 of 192 wells seeded with single DN1 cells showed growth of progeny cells (Fig. 2b, right panel). Of these 136 wells, 13 contained both T cells and macrophages. Macrophages were seen as large GFP$^+$ cells in these cultures (Fig. 2c). Green cells with dendritic morphology were also seen in almost all macrophage-T wells (Fig. 2d). Because early thymic progenitors are able to produce dendritic cells[21,22], it is important to distinguish macrophages from dendritic cells. A majority of large GFP$^+$ cells in these cultures contained phagosomal vacuoles (Supplementary Fig. 4a), and more than half of these cells were able to engulf latex beads (Fig. 2e), in a manner reminiscent of macrophages. Flow cytometric analysis confirmed the

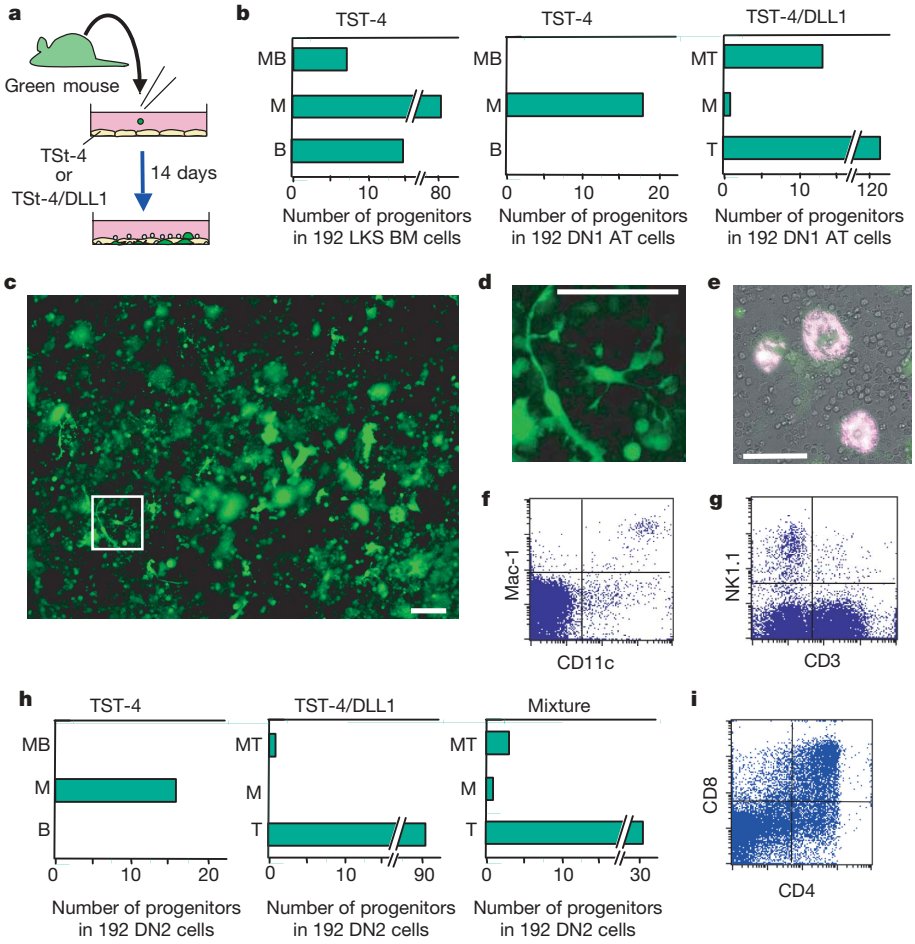


Figure 2 | T-cell progenitors in adult thymus that have lost B-cell potential retain macrophage potential. **a**, Experimental procedure for clonal analysis. Single cells were cultured for 14 days with the stromal cell line TSt-4 or TSt-4/DLL1 in the presence of stem cell factor, interleukin-7 and granulocyte/macrophage colony-stimulating factor. **b**, A total of 192 individual LKS bone marrow (BM) cells (left panel) and 192 DN1 adult thymus (AT) cells (middle panel) were cultured with TSt-4. Individual DN1 AT cells (total 192 cells) were also cultured with TSt-4/DLL1 (right panel). Seeded cells were classified according to the cells generated in each clonal culture: progenitors that generated both macrophages and B cells (MB), both macrophages and T (MT), only macrophages (M), only B cells (B) and only T cells (T). **c**, Photomicrograph of cells in a macrophage-T clone in the TSt-4/DLL1 cultures. Macrophages are seen as large GFP$^+$ cells. Most of the T-lineage cells are GFP$^-$ and are therefore not visible here. In this particular well, the total number of large GFP$^+$ cells (diameter more than 20 μm) was about 100, whereas that of T-lineage cells was about 10^5. Scale bar, 100 μm. **d**, Enlarged image of the area indicated in **c**. Scale bar, 100 μm. **e**, Photomicrograph taken 6 h after the addition of latex beads (1 μm diameter) to cells in a macrophage-T clone. Scale bar, 100 μm. **f**, **g**, Flow cytometric profiles of the cells recovered from a macrophage-T clone. Macrophages express Mac-1 and CD11c (**f**). NK-cell generation was determined by the appearance of CD3$^-$NK1.1$^+$ cells (**g**). **h**, A total of 192 DN2 AT cells were individually cultured with TSt-4 (left panel), with TSt-4/DLL1 (middle panel), or with a mixture of TSt-4 and TSt-4/DLL1 (4:1 ratio) (right panel). **i**, Flow cytometric profile of cells in a macrophage-T clone from a single DN2 cell shown in the right panel of **h**.

presence of Mac-1$^+$ cells in macrophage-T clonal cultures (Fig. 2f). Although these Mac-1$^+$ cells also expressed CD11c, a commonly used marker for dendritic cells, we consider them to be macrophages because they expressed F4/80 highly and showed efficient phagocytic activity (Supplementary Fig. 4c, d). In addition, the Mac-1$^+$CD11c$^+$ cells were much larger than the Mac-1^{low}CD11c$^+$ cells (Supplementary Fig. 4e). The macrophage–T-cell containing wells always contained NK cells (Fig. 2g). Similar results were obtained with fetal thymus progenitors (Supplementary Fig. 5).

We next examined the macrophage potential of DN2 cells, which have been shown to have completely lost B-cell potential, by culturing them with TSt-4 cells. The frequency of macrophage-generating progenitors was nearly comparable to that of the DN1 cells (16 of 192) (Fig. 2h, left panel). By using a mixture of TSt-4 and TSt-4/DLL1cells at a 4:1 ratio, 36 wells showed growth of progeny cells; of these, three showed growth of both T cells and macrophages (Fig. 2h, right panel, Fig. 2i and Supplementary Fig. 6). These results demonstrate that T-cell progenitors retain substantial macrophage potential after shutting off B-cell potential.

We then investigated whether thymic T-cell progenitors can generate macrophages within the thymic environment. Deoxyguanosine-treated fetal thymus lobes, which had been seeded with 200 DN2 cells from a green mouse several days earlier, were transplanted under the kidney capsule of a human (h) CD3εTg mouse[23] (Supplementary Fig. 7a), which completely lacks T-lineage cells as a result of a cell-autonomous defect in early T-cell progenitors[24]. All grafts showed a well-developed thymic architecture (Supplementary Fig. 7b, upper left panel). Macrophages, revealed by staining with anti-F4/80 antibody, were found mainly in the cortex of the graft (Supplementary Fig. 7b, lower left panel), a distribution also seen in the normal thymus[25]. A substantial number of F4/80$^+$ macrophages (about 20%) were GFP$^+$. Such GFP$^+$ macrophages were not seen in control grafts seeded with double-positive thymocytes from green mice (Supplementary Fig. 7c). In a serial section of the thymic lobe reconstituted with DN2 cells, a significant proportion of CD11c$^+$ cells were found to express GFP (Supplementary Fig. 7d), indicating that DN2 cells also can generate dendritic cells in the thymus.

Macrophage–T bipotentiality was further examined at a single-cell level with the same system as described above (Fig. 3a). Fifty adult DN1 cells were tested; reconstitution of thymocytes was observed in 20 lobes. Of these, five lobes showed completely restored thymic organization (Supplementary Fig. 8a, b). A proportion of the F4/80$^+$ cells in these grafts expressed GFP (Fig. 3b), providing clonal evidence that T-cell progenitors are able to give rise to thymic macrophages. In addition, a fraction of the CD11c$^+$ dendritic cells was also GFP$^+$ (Supplementary Fig. 8c).

We next injected DN2 adult thymus cells (1,000 cells) from green mice directly into the thymus of non-irradiated wild-type recipient mice (Fig. 3c). In thymic lobes harvested two weeks later, GFP$^+$ cells were seen mainly in medullary regions (Fig. 3d, upper right panel). GFP$^+$ macrophages were detected in the cortex close to these medullary regions (Fig. 3d, lower left panel).

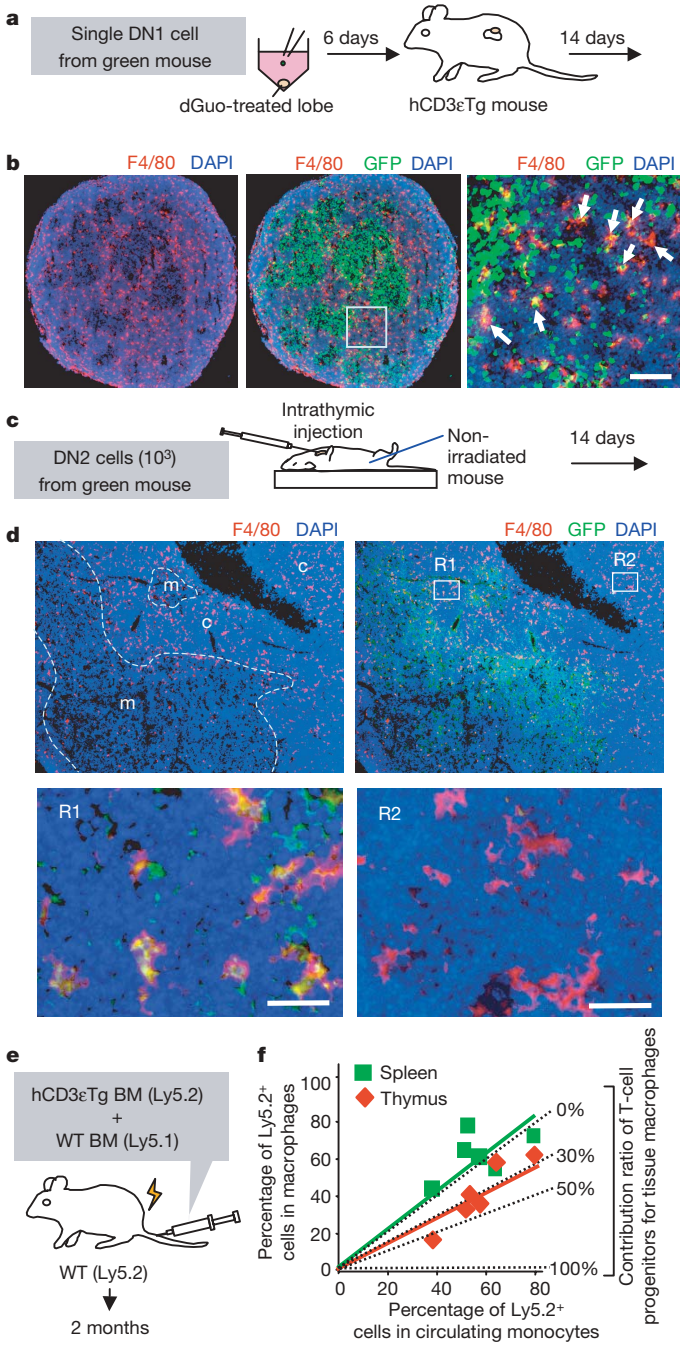


Figure 3 | *In vivo* evidence that T-cell progenitors contribute to the production of thymic macrophages. **a**, DN1 adult thymus cells from a green mouse were individually seeded onto deoxyguanosine-treated lobes and cultured for six days; the lobes were transplanted under the kidney capsule of adult hCD3εTg mice. The mice were killed two weeks later. **b**, A section of the graft was stained with F4/80 monoclonal antibody (red) and counterstained with 4,6-diamidino-2-phenylindole (DAPI; blue). The right panel shows an enlarged image for the gated area in the middle panel. Scale bar, 100 μm. Arrows indicate F4/80$^+$GFP$^+$ cells. **c**, DN2 adult thymus cells (1,000 cells) from a green mouse were injected directly into the thymus of non-irradiated wild-type mice and the thymuses were harvested 14 days later. **d**, A thymic section was stained with anti-F4/80 monoclonal antibody (red) and counterstained with DAPI (blue). The lower panels represent enlarged images of the areas indicated in the upper right panel as R1 (cortical region that is close to the medulla, which contains GFP$^+$ progeny of injected cells) and R2 (control cortical region). Scale bars, 50 μm. **e**, Experimental procedure to examine the contribution of T-cell progenitors to the production of thymic macrophages. A 1:1 mixture (5×10^6 for each) of wild-type bone marrow (BM) cells (Ly5.1) or bone marrow cells from hCD3εTg mice (Ly5.2) were injected intravenously into the tail vein of lethally irradiated wild-type Ly5.2 mice. The chimaeric mice were killed for analysis two months later. **f**, Derivation of a substantial proportion of thymic macrophages from thymic T-cell progenitors. Percentages of Ly5.2$^+$ cells in circulating monocytes (x axis) were determined by flow cytometry. The percentage of Ly5.2$^+$ cells among thymic and splenic F4/80$^+$ macrophages was determined by counting the cells in immunohistocheminally stained sections (y axis). Data from six chimaeric mice are plotted, and the line for splenic macrophages versus circulating monocytes (green line) and that for thymic macrophages versus circulating monocytes (red line) were fitted by least-squares analysis. The dotted lines represent the predicted outcome assuming that the contribution of T-cell progenitors to macrophage production in the tissue is 0%, 30%, 50% or 100%.

Because the cortex is a site where DN1 and DN2 cells expand[26], it is likely that these cells produce cortical macrophages in addition to T cells. In contrast, just as in other tissues, circulating monocytes may also be a source of thymic macrophages. To estimate the extent that thymic T-cell progenitors contribute to the production of thymic macrophages, we transplanted a 1:1 mixture of wild-type bone marrow cells (Ly5.1) and bone marrow cells from hCD3εTg mice (Ly5.2) into lethally irradiated wild-type Ly5.2 mice (Fig. 3e). Two months later, all thymic T cells were $Ly5.1^+$ as expected, because the $Ly5.2^+$ hCD3εTg bone marrow cells cannot give rise to T-lineage cells, including DN1 and DN2 cells (data not shown). The proportion of $Ly5.2^+$ cells among circulating monocytes fluctuated between recipients and ranged from 40% to 80%. If thymic macrophages are derived exclusively from circulating monocytes, the proportion of $Ly5.2^+$ thymic macrophages should be the same as that of the circulating $Ly5.2^+$ monocytes. In contrast, if thymic macrophages are produced exclusively from double-negative thymocytes, all thymic macrophages should be $Ly5.1^+$. The chimaerism ratios from individual mice for splenic macrophages versus circulating monocytes were plotted (green rectangles in Fig. 3f), and the derived line (green line) through the data points fell along the line representing the identical chimaerism ratio, indicating that almost all splenic macrophages are derived from circulating monocytes. By contrast, the line derived from the data points for the chimaerism ratio for thymic macrophages versus circulating monocytes (red line) had a smaller slope than that of the splenic macrophage line. From the inclination of the line it can be predicted that about 30% of thymic macrophages are derived from thymic progenitors that also can generate T cells.

We have shown here that a large majority of myeloid cells derived from thymic progenitors were macrophages. However, in the accompanying paper[27] the authors show that thymic progenitors produce granulocytes in addition to macrophages, in experiments both *in vitro* and *in vivo*. The inconsistency with the *in vitro* data could be due to a difference in the culture period, because a longer period is required for macrophages to develop in significant numbers. Moreover, flow cytometric analysis used in their *in vivo* studies could detect granulocytes more efficiently than macrophages, whereas the reverse is true for the histological analysis we used.

The present study therefore provides evidence, at a clonal level, for the existence of thymic T-cell progenitors that have lost B-cell potential but still retain macrophage potential (Fig. 4). We show further that this macrophage potential is not merely a phenomenon *in vitro* but occurs *in vivo* in the thymus. Our present findings indicate that the progenitor cell at the branch point of the T-cell and B-cell lineages retains macrophage potential, thus formally proving that the CLP stage is dispensable in the developmental pathway from haematopoietic stem cells to T cells. We propose that the myeloid-based model holds true for both fetal and adult haematopoiesis (Supplementary Fig. 1b). We emphasize that the fetal haematopoiesis described here is not the primitive haematopoiesis that takes place in the yolk sac but belongs in the same category as adult haematopoiesis, which is referred to as 'definitive' haematopoiesis, producing adult-type blood cells in fetal liver and subsequently in bone marrow. Because lineage restriction programmes in haematopoiesis should have been constructed step by step during evolution, it seems logical to assume that the same or at least a similar programme is used regardless of the anatomical site where haematopoiesis occurs.

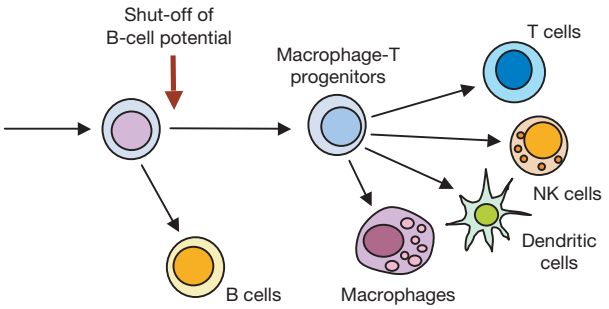


Figure 4 | T-cell progenitors retain macrophage potential after B-cell potential has been shut off. The present study shows that early T-cell progenitors in the adult thymus that have lost B-cell potential still retain a substantial ability to generate macrophages. It is therefore unlikely that the CLP stage exists along the normal physiological pathway towards T cells in adult haematopoiesis. Our present findings argue against the classic model of haematopoiesis in which the CLP is located at the branch point towards T and B-cell lineages, and strongly suggest that the myeloid-based model is applicable to both fetal and adult haematopoiesis.

METHODS SUMMARY

Transgenic mice carrying an enhanced GFP gene ('green mice')[20] were donated by M. Okabe. hCD3εTg mice[23] were donated by C. Terhorst. To assess the developmental potential of progenitors *in vitro*, co-culture with different stromal cells was mainly used. For the *in vivo* experiments, the following methods were used: transplantation of deoxyguanosine-treated fetal thymus lobes, seeded with progenitors to be tested, under the kidney capsule of hCD3εTg mice; intrathymic injection; and an *in vivo* competition assay in which a 1:1 mixture of bone marrow cells from B6 Ly5.1 mice and from hCD3εTg mice (Ly5.2) were injected intravenously into lethally irradiated wild-type Ly5.2 mice. The proportion of $Ly5.2^+$ cells among $F4/80^+$ cells was determined on immunohistochemically stained thymic sections from the chimaeric mice.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 October 2007; accepted 15 February 2008.

1. Kawamoto, H., Ohmura, K. & Katsura, Y. Direct evidence for the commitment of hematopoietic stem cells to T, B and myeloid lineages in murine fetal liver. *Int. Immunol.* **9,** 1011–1019 (1997).
2. Kawamoto, H., Ohmura, K. & Katsura, Y. Presence of progenitors restricted to T, B, or myeloid lineage, but absence of multipotent stem cells, in the murine fetal thymus. *J. Immunol.* **161,** 3799–3802 (1998).
3. Ohmura, K. *et al.* Emergence of T, B and myeloid lineage-committed as well as multipotent hematopoietic progenitors in the aorta–gonad–mesonephros region of day 10 fetuses of the mouse. *J. Immunol.* **163,** 4788–4795 (1999).
4. Kawamoto, H., Ikawa, T., Ohmura, K., Fujimoto, S. & Katsura, Y. T cell progenitors emerge earlier than B cell progenitors in the murine fetal liver. *Immunity* **12,** 441–450 (2000).
5. Ikawa, T. *et al.* Identification of the earliest prethymic T-cell progenitors in murine fetal blood. *Blood* **103,** 530–537 (2004).
6. Lu, M., Kawamoto, H., Katsube, Y., Ikawa, T. & Katsura, Y. The common myelolymphoid progenitor: a key intermediate stage in hemopoiesis generating T and B cells. *J. Immunol.* **169,** 3519–3525 (2002).
7. Katsura, Y. Redefinition of lymphoid progenitors. *Nature Rev. Immunol.* **2,** 127–132 (2002).
8. Kawamoto, H. A close developmental relationship between the lymphoid and myeloid lineages. *Trends Immunol.* **27,** 169–175 (2006).
9. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91,** 661–672 (1997).
10. Kincade, P. W. *et al.* Nature or nurture? Steady-state lymphocyte formation in adults does not recapitulate ontogeny. *Immunol. Rev.* **187,** 116–125 (2002).
11. Spangrude, G. J. Divergent models of lymphoid lineage specification: do clonal assays provide all the answers? *Immunol. Rev.* **187,** 40–47 (2002).
12. Laiosa, C. V., Stadtfeld, M. & Graf, T. Determinants of lymphoid–myeloid lineage diversification. *Annu. Rev. Immunol.* **24,** 705–738 (2006).
13. Buza-Vidas, N., Luc, S. & Jacobsen, S. E. Delineation of the earliest lineage commitment steps of haematopoietic stem cells: new developments, controversies and major challenges. *Curr. Opin. Hematol.* **14,** 315–321 (2007).
14. Lee, C. K. *et al.* Generation of macrophages from early T progenitors *in vitro*. *J. Immunol.* **166,** 5964–5969 (2001).
15. Balciunaite, G., Ceredig, R. & Rolink, A. G. The earliest subpopulation of mouse thymocytes contains potent T, significant macrophage, and natural killer cell but no B-lymphocyte potential. *Blood* **105,** 1930–1936 (2005).
16. Benz, C. & Bleul, C. C. A multipotent precursor in the thymus maps to the branching point of the T versus B lineage decision. *J. Exp. Med.* **202,** 21–31 (2005).
17. Nishikawa, S., Ogawa, M., Nishikawa, S., Kunisada, T. & Kodama, H. B lymphopoiesis on stromal cell clone: stromal cell clones acting on different stages of B cell differentiation. *Eur. J. Immunol.* **18,** 1767–1771 (1988).
18. Anderson, M. K., Weiss, A. H., Hernandez-Hoyos, G., Dionne, C. J. & Rothenberg, E. V. Constitutive expression of PU.1 in fetal hematopoietic progenitors blocks T cell development at the pro-T cell stage. *Immunity* **16,** 285–296 (2002).
19. Masuda, K. *et al.* Prethymic T-cell development defined by the expression of paired immunoglobulin-like receptors. *EMBO J.* **24,** 4052–4060 (2005).

20. Ikawa, M., Yamada, S., Nakanishi, T. & Okabe, M. 'Green mice' and their potential usage in biological research. *FEBS Lett.* **430,** 83–87 (1998).
21. Ardavin, C., Wu, L., Li, C. L. & Shortman, K. Thymic dendritic cell and T cells develop simultaneously in the thymus from a common precursor population. *Nature* **362,** 761–763 (1993).
22. Lu, M. *et al.* The earliest thymic progenitors in adults are restricted to T, NK, and dendritic cell lineage and have a potential to form more diverse TCRβ chains than fetal progenitors. *J. Immunol.* **175,** 5848–5856 (2005).
23. Wang, B. *et al.* A block in both early T lymphocyte and natural killer cell development in transgenic mice with high-copy numbers of the human CD3E gene. *Proc. Natl Acad. Sci. USA* **91,** 9402–9406 (1994).
24. Hollander, G. A. *et al.* Developmental control point in induction of thymic cortex regulated by a subpopulation of prothymocytes. *Nature* **373,** 350–353 (1995).
25. Surh, C. D. & Sprent, J. T-cell apoptosis detected in situ during positive and negative selection in the thymus. *Nature* **372,** 100–103 (1994).
26. Lind, E. F., Prockop, S. E., Porritt, H. E. & Petrie, H. T. Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development. *J. Exp. Med.* **194,** 127–134 (2001).
27. Bell, J. J. & Bhandoola, A. The earliest thymic progenitors for T cells possess myeloid lineage potential. *Nature* doi:10.1038/nature06840 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Burrows, W. van Ewijk and W. T. V. Germeraad for critical reading of the manuscript, and O. Kawamoto for valuable advise.

Author Contributions H.W. performed most of experiments, with the assistance of R.S. and K.K. for immunohistochemical analysis and stromal cell co-cultures, respectively. Experiments shown in Fig. 1 and Supplementary Figs 2 and 3 were performed by K.M., and those in Supplementary Fig. 4d by T.I. Y.K. gave critical advice and comments in designing the experiments and writing the paper. H.K. designed the experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.K. (kawamoto@rcai.riken.jp).

METHODS

Mice. C57BL/6 (B6), B6Ly5.1 and Balb/c mice were purchased from CLEA Japan Inc. Transgenic mice on an B6 background carrying an enhanced GFP gene (EGFP Tg mice)[20] were donated by M. Okabe. hCD3εTg mice[23] were donated by C. Terhorst. Embryos at various stages of gestation were obtained from timed pregnancies. The day of observing the vaginal plug was designated as 0 days after coitus.

Antibodies. The following antibodies were purchased from BD Pharmingen: Ly5.1 (A20), Ly5.2 (104), c-Kit (2B8), Sca-1 (D7), erythroid lineage cells (TER119), Mac-1 (M1/70), Gr-1 (RB6-8C5), CD11c (HL3), B220 (RA3-6B2), Thy 1.2 (53-2.1), CD8 (53-6.7), CD4 (H129.19), NK1.1 (PK136), CD3ε (145-2C11), CD19 (1D3), anti-class II (M5/114), F4/80 (BM8), and CD25 (PC61). TER119, Mac-1, Gr-1, B220, CD19, NK1.1, CD3ε, CD4 and CD8 were used as Lin markers. For immunohistochemistry, the following antibodies were used: F4/80, biotinylated CD11c and biotinylated mouse anti-K8 (PROGEN) and ER-TR5 (donated by W. van Ewijk) as primary antibody, followed by goat anti-rabbit IgG-Texas Red conjugate, Alexa Fluor546 goat anti-rat IgG (H+L) conjugate or Alexa Fluor546 streptavidin conjugate (Molecular Probes), as secondary reagents.

Preparation of cells from progenitor populations. Single-cell suspensions of adult thymus cells were prepared from 8–12-week-old B6 mice as described previously[22]. Cells were then treated for 30 min at 37 °C with rabbit complement in the presence of anti-CD8 (3.155) and anti-CD3 (145-2C11) monoclonal antibody. The cells surviving the complement killing were used for cell sorting. The procedure for the preparation of progenitor enriched populations from fetal thymus and fetal liver has been described elsewhere[19]. Bone marrow cells were obtained from femurs of ten-week-old adult mice.

Growth factors. Recombinant murine (rm) stem cell factor (SCF), interleukin (IL)-1α, IL-3, IL-7, Flt-3L, granulocyte colony-stimulating factor (G-CSF), macrophage colony-stimulating factor (M-CSF) and granulocyte/macrophage colony-stimulating factor (GM-CSF) were purchased from R&D Systems.

Co-culture with stromal cells. To assess granulocyte/macrophage potential, 100 cells per well were cultured with PA6 (ref. 17) stromal cells in the presence of granulocyte colony-stimulating factor (10 ng ml^{-1}) for 7 days. Generated cells were counted and then stained with Mac-1; Mac-1^{+} cells sorted by fluorescence-activated cell sorting were centrifuged onto glass slides for Wright staining. For the detection of the B-cell and myeloid potential of progenitors, single cells from green mice were cultured for 14 days with TSt-4 cells in 96-well plates. Generation of macrophages was determined by fluorescence microscopy, and that of B cells was detected by subsequent flow cytometric analysis of CD19 expression. For the detection of T-cell and macrophage potential, TSt-4 cells that had been retrovirally transduced with the murine *Dll-1* gene (TSt-4/DLL1 cells)[19] were used. Medium was supplemented with SCF (1 ng ml^{-1}), IL-7 (0.5 ng ml^{-1}), Flt-3L (5 ng ml^{-1}) and GM-CSF (1 ng ml^{-1}). Generation of macrophages was determined as above, and that of T cells was detected by subsequent flow cytometric analysis. All co-cultures were maintained in RPMI 1640 medium (Sigma) supplemented with 10% (v/v) FCS, L-glutamine (2 mM), sodium pyruvate (1 mM), sodium bicarbonate (2 mg ml^{-1}), non-essential amino acid solution (0.1 mM) (Gibco BRL), 2-mercaptoethanol (50 μM), streptomycin (100 mg ml^{-1}), and penicillin (100 U ml^{-1}).

CFU-C assay. Cells (1,000 cells per dish) were cultured for seven days in triplicate dishes in α-minimal essential medium (GibcoBRL) containing 30% FCS, 1% methylcellulose, 1% bovine serum albumin, 2-mercaptoethanol (50 μM), L-glutamine (1 mM), SCF (10 ng ml^{-1}), IL-3 (10 ng ml^{-1}), GM-CSF (10 ng ml^{-1}), IL-1α (10 ng ml^{-1}), G-CSF (10 ng ml^{-1}) and M-CSF (10 ng ml^{-1}).

Phagocytosis assay. Carboxylate-modified microspheres (FluoSpheres, red, 1 μm; Molecular Probes) were added to the wells that were judged to contain macrophages. After 6 h, cells were gently washed with culture medium. All large GFP^{+} cells (more than 20 μm in diameter) with or without engulfed beads in each clonal culture were counted by fluorescence-microscopic observation. In the experiment shown in Supplementary Fig. 4d, sorted cells were incubated with Fluoresbrite YG beads (1 μm; Molecular Probes) for 6 h in the presence of M-CSF (10 ng ml^{-1}).

Fetal thymus organ culture and kidney subcapsular transplantation. The basic procedure of fetal thymus organ culture under high-oxygen-submersion conditions has been described previously[1,22]. In brief, single deoxyguanosine-treated fetal thymus lobes at 15 days after coitus were placed into wells of a 96-well V-bottomed plate. DN2 adult thymus cells (200 cells per lobe) or DN1 adult thymus cells (one cell per lobe) were cultured for six days with a deoxyguanosine-treated lobe under high-oxygen-submersion conditions in the presence of SCF (1 ng ml^{-1}), IL-7 (0.5 ng ml^{-1}), Flt-3L (5 ng ml^{-1}) and GM-CSF (1 ng ml^{-1}). The medium used for deoxyguanosine treatment and high-oxygen-submersion cultures was the same as used for co-culture with stromal cells. The lobes were then transplanted individually under the kidney capsule of recipient mice.

Intrathymus cell transfer experiment. Intrathymic injection of sorted thymocytes was performed as described previously[22]. DN2 adult thymus cells (1,000 cells) were injected into the thymic lobes of non-irradiated recipient mice.

Bone marrow transplantation experiments. In experiments shown in Fig. 3e, f, a 1:1 mixture of wild-type bone marrow cells (B6 Ly5.1 mice) (5×10^6 cells) and bone marrow cells from hCD3εTg mice (Ly5.2) (5×10^6 cells) was transplanted into lethally irradiated wild-type Ly5.2 mice. The resulting chimaeric mice were killed and analysed two months later. For assessing the chimaerism ratio in macrophages residing in the thymic cortex, the proportion of Ly5.2^{+} cells among F4/80^{+} cells was determined on immunohistochemically stained thymic sections from the chimaeric mice. To establish baseline values for these experiments, we reconstituted lethally irradiated wild-type Ly5.2 recipient mice with wild-type Ly5.1 bone marrow cells (10^7 cells). The proportions of Ly5.2^{+} cells were 36.3% (s.d. 19.1%; $n = 4$) and 31.9% (s.d. 10.6%; $n = 4$) among thymic and splenic macrophages, respectively. For the analysis in Fig. 3f, these proportions were regarded as 0%. The lines for the experimental data were fitted by least-squares analysis.

RT–PCR. RT–PCR was performed as described previously[4]. Primers used are described in the legend to Supplementary Fig. 3. PCR products were subjected to electrophoresis through a 1.2% agarose gel and stained with ethidium bromide.

Immunohistochemistry. Tissues were fixed for 15 min with 4% paraformaldehyde, placed in sucrose gradients, embedded in OCT compound, and snap-frozen in liquid nitrogen. Frozen blocks were cut into serial 5-μm sections and mounted onto MAS-coated microscope slides (Matsunami). Sections were incubated with primary antibodies and washed with PBS/0.01% Tween; this was followed by incubation with the appropriate secondary reagents. Nuclei were counterstained with DAPI (Molecular Probes).

LETTERS

Impaired T_H17 cell differentiation in subjects with autosomal dominant hyper-IgE syndrome

Joshua D. Milner[1]*, Jason M. Brenchley[2]*†, Arian Laurence[3], Alexandra F. Freeman[4], Brenna J. Hill[2], Kevin M. Elias[3,5], Yuka Kanno[3], Christine Spalding[4], Houda Z. Elloumi[4], Michelle L. Paulson[4], Joie Davis[4], Amy Hsu[4], Ava I. Asher[2], John O'Shea[3], Steven M. Holland[4], William E. Paul[1] & Daniel C. Douek[2]

The autosomal dominant hyper-IgE syndrome (HIES, 'Job's syndrome') is characterized by recurrent and often severe pulmonary infections, pneumatoceles, eczema, staphylococcal abscesses, mucocutaneous candidiasis, and abnormalities of bone and connective tissue[1,2]. Mutations presumed to underlie HIES have recently been identified in *stat3*, the gene encoding STAT3 (signal transducer and activator of transcription 3) (refs 3, 4). Although impaired production of interferon-γ and tumour-necrosis factor by T cells[5], diminished memory T-cell populations, decreased delayed-type-hypersensitivity responses and decreased *in vitro* lymphoproliferation in response to specific antigens[6] have variably been described, specific immunological abnormalities that can explain the unique susceptibility to particular infections seen in HIES have not yet been defined. Here we show that interleukin (IL)-17 production by T cells is absent in HIES individuals. We observed that *ex vivo* T cells from subjects with HIES failed to produce IL-17, but not IL-2, tumour-necrosis factor or interferon-γ, on mitogenic stimulation with staphylococcal enterotoxin B or on antigenic stimulation with *Candida albicans* or streptokinase. Purified naive T cells were unable to differentiate into IL-17-producing (T_H17) T helper cells *in vitro* and had lower expression of retinoid-related orphan receptor (ROR)-γt, which is consistent with a crucial role for STAT3 signalling in the generation of T_H17 cells[7–14]. T_H17 cells have emerged as an important subset of helper T cells[15] that are believed to be critical in the clearance of fungal[16] and extracellular bacterial[17] infections. Thus, our data suggest that the inability to produce T_H17 cells is a mechanism underlying the susceptibility to the recurrent infections commonly seen in HIES.

We studied three groups of subjects: healthy controls with no apparent immunological defects, HIES individuals with defined mutations in *stat3*, and individuals (termed 'HIES-like') with some combination of elevated IgE, atopic dermatitis, skeletal abnormalities and susceptibility to infection, but without recurrent staphylococcal abscesses or candidiasis or *stat3* mutations (Table 1).

We observed that IL-17-producing T cells were barely detectable among peripheral blood mononuclear cells (PBMCs) from subjects with HIES on stimulation with staphylococcal enterotoxin B (SEB) (Fig. 1a). The frequency of SEB-induced interferon (IFN)-γ-producing CD4 T cells from PBMCs of subjects with HIES was similar to that of healthy controls, whereas the frequency of cells producing IL-2 and/or tumour-necrosis factor (TNF) was slightly reduced. Fewer of the IFN-γ-producing CD4 T cells from subjects with HIES also produced TNF and/or IL-2 compared with healthy controls (Fig. 1b). IL-17-producing T cells were present in PBMCs from the HIES-like cohort with no mutations in *stat3*, suggesting that the lack of IL-17 production in HIES has a critical function in susceptibility to the specific infections seen in HIES, and also that elevated serum IgE, atopic dermatitis or low frequency of memory T cells (data not shown) are not independently associated with severe defects in the T_H17 axis (Fig. 1a). The production of IL-21 and IL-22, which have been described as both T_H1 and T_H17 cytokines[9,10,18], was not significantly lower in subjects with HIES than in healthy controls. Additionally, among subjects with HIES, IL-21 was derived predominantly from T_H1 cells, suggesting that the induction of IL-21 and IL-22 is not uniquely dependent on normal STAT3 signalling in humans (Supplementary Fig. 1a–c). Although the HIES-like subjects had significantly fewer IFN-γ-producing cells than healthy controls did (Fig. 1a), they are not susceptible to the specific infections seen in HIES, namely recurrent candidiasis and staphylococcal abscesses.

We next tested whether T_H17 cells could be generated from CD4 T cells of subjects with HIES. Naive CD4 T cells ($CD4^+CD8^-CD45RO^-CD11a^{dull}CD27^+CD31^+$) from healthy controls and subjects with HIES were sorted by flow cytometry and then cultured in T_H17 conditions[7,10] with anti-CD2/CD3/CD28 beads in the presence of anti-IFN-γ alone, or with the addition of either IL-6 and IL-1β, or IL-23 and IL-1β, for 12 days. T_H17 cell differentiation was observed in all CD4 T-cell cultures from healthy controls. However, CD4 T cells from the majority of subjects with HIES showed variable survival when cultured with either IL-6 or IL-23 in the presence of IL-1β, but not when cultured without additional cytokines. Surviving cells did not make any IL-17, but IFN-γ-producing CD4 T cells were present, although at a lower frequency than among CD4 T cells from controls (Fig. 2a–e). *ROR-γt* mRNA expression in HIES T cells after 48 h under T_H17-polarizing conditions was significantly lower in T cells from subjects with HIES than in healthy controls (Fig. 2f). Introduction of mutant *stat3* from subjects with HIES, with either SH2 domain or DNA-binding mutations, into normal human naive CD4 T cells also resulted in a decrease in *ROR-γt* mRNA expression under the same T_H17-polarizing conditions (25% decrease for the SH2 domain mutation, and 65% for the DNA-binding mutation), showing that mutant STAT3 suppresses ROR-γt expression in a dominant-negative fashion (Supplementary Fig. 1d). Given the variable cell survival in the T_H17 differentiation experiments, we directly tested the effects of IL-1β on T-cell survival by culturing PBMCs from subjects with HIES and from healthy individuals overnight, and measuring subsequent IL-1β-induced apoptosis by staining with 7-amino-actinomycin D (7-AAD) and 3,3′-dihexyloxacarbocyanine

[1]Laboratory of Immunology, [2]Human Immunology Section, Vaccine Research Center, and [3]Lymphocyte Cell Biology Section, Molecular Immunology and Inflammation Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, [4]Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [5]Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. †Present address: Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

Table 1 | Summary of patient characteristics

Patient ID	Age (yr)	Sex	*stat3* mutation	HIES clinical score	Serum IgE* (IU ml^{-1})	Skeletal abnormalities	Recurrent candidiasis/staphylococcal abscess or lung infection	Atopic dermatitis
1† (J112)	6	M	1970 A→G (SH2)	31	4,190	+	+	+
2 (J088)	7	M	1145 G→T (DNA)	70	18,000	+	+	+
3 (J083)	10	F	1909 G→A (SH2)	78	5,070	+	+	+
4 (J005)	13	F	1144 C→T (DNA)	60	18,600	+	+	+
5 (J100)	22	F	1909 G→A (SH2)	76	1,020	+	+	+
6 (J017)	23	M	1145 G→A (DNA)	82	20,500	+	+	+
7‡ (J002)	24	F	1865 C→T (SH2)	85	8,550	+	+	+
8† (J112)	51	M	1970 A→G (SH2)	79	6,380	+	+	+
9 (J054)	40	F	1393 T→G (DNA)	78	239 (47,338)	+	+	+
10‡ (J002)	56	M	1865 C→T (SH2)	90	26 (2,392)	+	+	+
11	48	M	1268 G→A (DNA)	65	1,340 (20,700)	+	+	+
12	37	F	1909 G→A	94	495 (25,058)	+	+	+
13 (J015)	36	F	1861 T→G	96		+	+	+
14§	16	M	None	27	11,100	−	− (other recurrent infections)	−
15§	13	F	None	30	14,000	−	− (other recurrent infections)	−
16§	18	F	None	n.a.	136	−	− (other recurrent infections)	−
17	15	M	None	49	6,880	+	−	+
18	15	M	None	55	160 (69,280)	+	−	+
19	4	F	None	n.d.	>30,000	+	−	+

n.a., not available; n.d., not determined. *Peak values in parentheses. †‡Parent–child pairs. §Siblings with similar phenotypes.

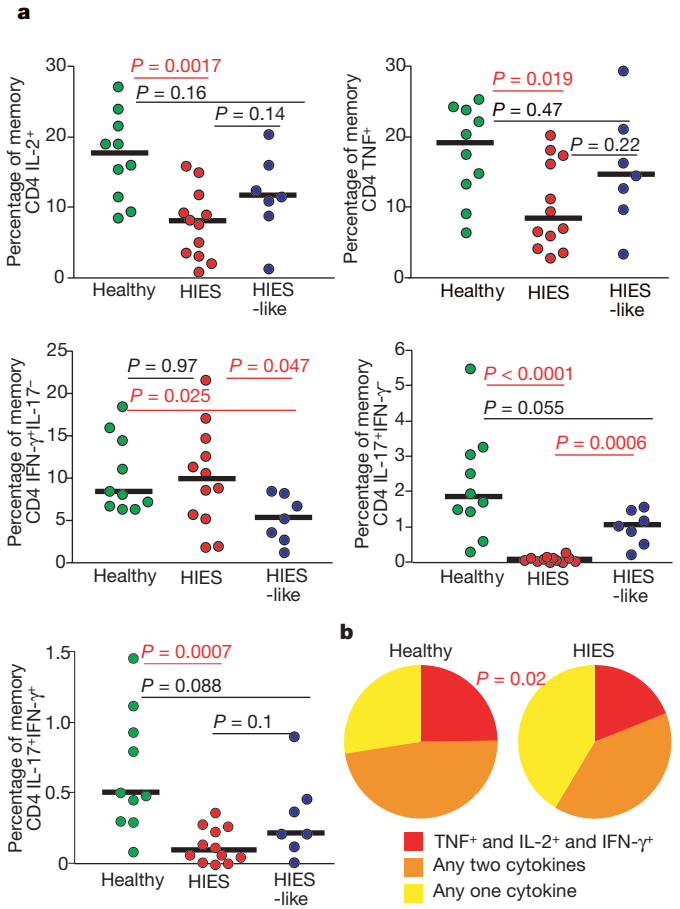


Figure 1 | Lack of IL-17 production in SEB-stimulated PBMCs from HIES patients. a, PBMCs from cohorts of HIES patients with *stat3* mutations (red circles), HIES-like patients (blue circles) and healthy individuals (green circles) were stimulated overnight with SEB in the presence of brefeldin A and stained as described in Methods. The frequency of memory CD4 T cells (gated as described in Methods) that produced IFN-γ, IL-2, TNF or IL-17 was then assessed. **b**, Memory CD4 T cells that produced IFN-γ were analysed for their ability to co-produce IL-2 and TNF by using SPICE (Simplified Presentation of Incredibly Complex Evaluations) as described in Methods. Statistical significance was determined with the Mann–Whitney test. Significant *P* values are shown in red, and median values with horizontal bars.

iodide (DiOC6). Apoptosis was significantly increased among CD4 T cells from subjects with HIES in comparison with healthy individuals or HIES-like individuals in the presence of IL-1β (Supplementary Fig. 2a). We observed no differences between controls and subjects with HIES in the percentage of FoxP3$^+$ cells in sorted naive CD4 T cells or among total CD4 T cells (data not shown). Furthermore, IL-4 production was not enhanced in stimulated purified naive CD4 T cells from subjects with HIES; even after one round of priming under T_H2-polarizing conditions, IL-4 production was undetectable in either subjects with HIES or controls. In addition, one round of priming in T_H1-polarizing conditions or conditions with no added cytokines led to similar frequencies of IFN-γ-producing cells in subjects with HIES and in controls (Supplementary Fig. 3).

The failure of HIES T cells to differentiate into T_H17 cells is likely to be due to mutated *stat3*; however, aberrant STAT3 DNA binding in primary T cells from HIES patients has not previously been reported. We therefore performed an electrophoretic mobility-shift assay with purified HIES T cells treated with the potent *stat3*-activating and T_H17-promoting cytokine IL-21. We found that, even in the presence of normal levels of phosphorylated STAT3, binding of STAT3 to DNA was undetectable, whereas robust binding was found in similarly treated purified T cells from healthy donors (Supplementary Fig. 4). Therefore, although IL-21 production seems normal in HIES, its signalling is not.

Although cells capable of producing IL-2, TNF and IFN-γ were present in the memory CD4 T-cell population from subjects with HIES (Fig. 1), it remained possible that the lack of T_H17 cells in the affected individuals was due to a more global failure of T-cell maturation and differentiation. Indeed, consistent with previous studies[6] was our observation that the frequency of memory CD4 and CD8 T cells, defined by the expression of CD45RO, was significantly decreased in subjects with HIES, and in many of the HIES-like subjects (data not shown). PBMCs were stimulated with streptokinase (a streptococcal antigen), *Candida albicans* (CA) or *Staphylococcus aureus* (SA), all antigens from microorganisms commonly associated with the infections that characterize HIES. The percentage of memory CD4 T cells producing IL-2, TNF and IFN-γ in response to CA or streptokinase was not decreased in HIES-like subjects or in subjects with HIES in comparison with healthy controls. However, HIES T cells showed little or no antigen-specific IL-17 production (Fig. 3a–c), whereas both healthy controls and subjects with HIES were able to mount antigen-specific IL-17 responses. There was significant lymphocyte cell death in the cultures from patients with HIES that had been incubated with SA (Supplementary Fig. 2b). Although SEB stimulation did not cause cell death, other staphylococcal T-cell mitogens or innate immune ligands might be responsible. In a previous study[19] a

microarray analysis of purified T cells from subjects with HIES stimulated with SA showed diminished IL-17 mRNA expression, providing further evidence that pathogen-induced IL-17 production is impaired in subjects with HIES.

HIES is a multifaceted disease, a major component of which is immunodeficiency. Herein we describe a potential mechanism for the susceptibility to particular fungal and extracellular bacterial infections, namely the inability to generate T_H17 cells. Mice with a targeted mutation in IL-17 are susceptible to *C. albicans* and the extracellular bacterium *Klebsiella pneumoniae*[16,17], presumably as a result of abnormal antimicrobial peptide expression[20] and neutrophil recruitment[21], mobilization and influx[16]. This suggests that it is the absence of T_H17 cells themselves, and not other STAT3-dependent

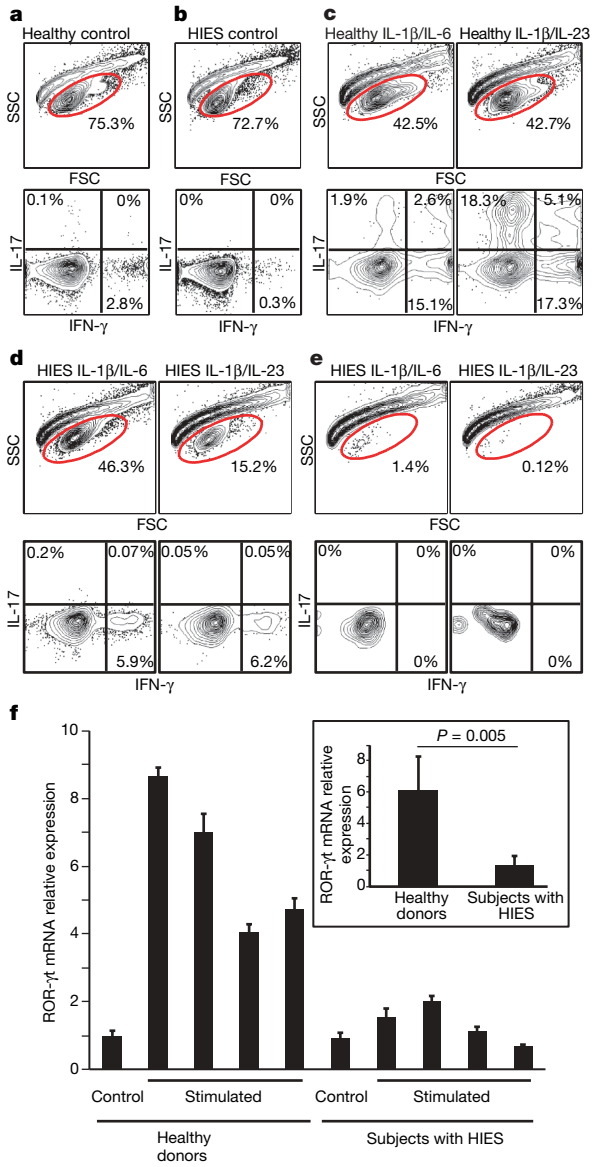


Figure 2 | Failure of T_H17 generation from naive cells of patients with HIES. A total of 50,000 naive T cells from healthy controls and patients with HIES were stimulated with anti-CD2/CD3/CD28 microbeads and the indicated cytokines as described in Methods. **a**–**e**, Dot plots of intracellular cytokine staining from healthy control (**a**) and HIES (**b**) cells stimulated without any cytokines, and from control subjects (**c**) and subjects with HIES (**d**, **e**) stimulated in the presence of IL-6 and IL-1β, or with IL-23 and IL-1β. **e**, Representative forward and side scatter plots of HIES cells after stimulation for 12 days show the poor recovery after stimulation with either IL-6 and IL-1β, or IL-23 and IL-1β. **d**, The one subject sample that had good cellular recovery. **f**, Naive CD4 T cells from four healthy donors and four subjects with HIES were freshly lysed or stimulated with anti-CD3 and anti-CD28 in the presence of IL-23 for 48 h. *ROR-γt* mRNA expression was detected by quantitative RT–PCR. Values shown are relative expression levels of triplicate samples (means and s.d.). Inset: average expression levels for the four healthy donors compared with those for the four subjects with HIES. Statistical significance was determined with a two-tailed unpaired Student *t*-test.

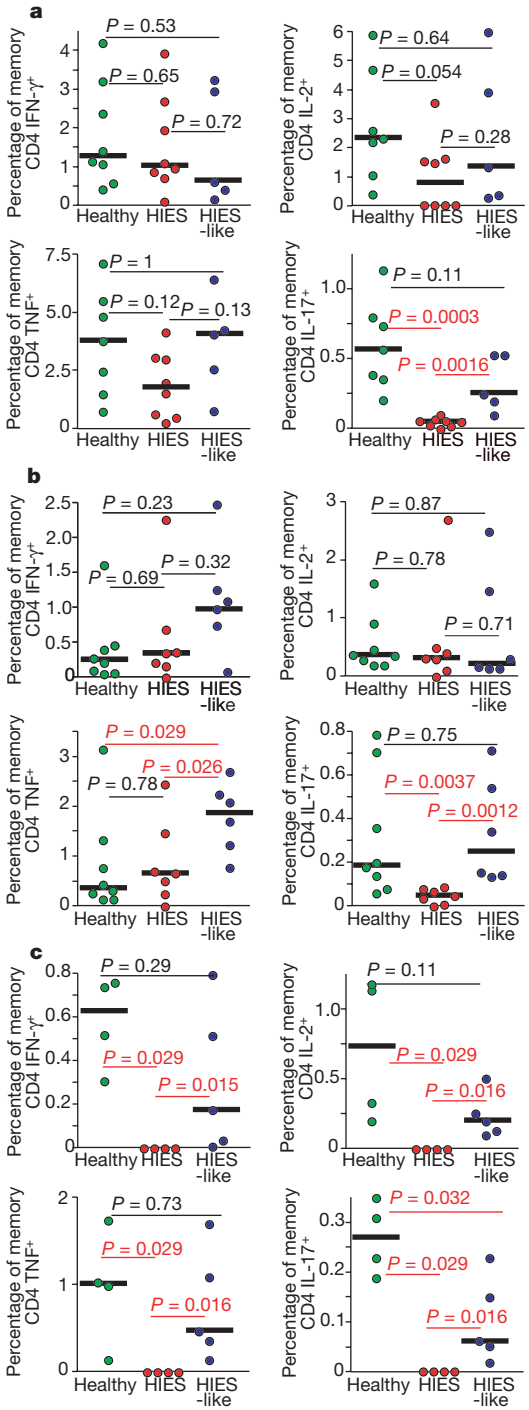


Figure 3 | Lack of IL-17 production despite antigen-specific IFN-γ, IL-2 and TNF production from subjects with HIES. PBMCs from cohorts of healthy subjects (green circles) HIES-like subjects (blue circles) and individuals with HIES (red circles) were stimulated overnight with *Candida albicans* (**a**), streptococcal kinase (**b**) or *Staphylococcus aureus* (**c**) in the presence of brefeldin A, then stained as described in Methods. The production of IFN-γ, IL-2, TNF or IL-17 by memory CD4 T cells (gated as described in Methods) in response to individual antigen stimulation is shown. Statistical significance was determined with the Mann–Whitney test. Significant *P* values are shown in red, and median values with horizontal bars.

mechanisms, that render subjects with HIES susceptible to these infections. The cell death we have observed *in vitro* in response to *S. aureus* may be responsible for the lack of cytokine responses to SA *ex vivo* (Fig. 3c), and this may have implications for the control of *S. aureus* infection *in vivo*. Further examination of mouse models of the dominant-negative HIES *stat3* mutations should help test the role of IL-17 in anti-staphylococcal responses more directly. IL-1β, which also induced cell death in subjects with HIES, is not only crucial for the generation of T_H17 cells, it is also critical to the clearance of *S. aureus* in mice through neutrophil recruitment[22].

Lack of IL-17 may also function in the eosinophilia of HIES, because neutralization of IL-17 on allergenic challenge of sensitized mice leads to increased eosinophilic disease[23], whereas the elevated IgE levels in subjects with HIES are likely to have a complex basis. IL-21, which can signal through STAT1 and STAT5, as well as STAT3, could be involved, because one study of *IL-21R*$^{-/-}$ mice reported elevated IgE levels[24], although another did not[25]. Because IL-21 is critical for immunoglobulin production[26], and because non-IgE immunoglobulin levels are normal in HIES, one could speculate that the normal IL-21 production we have observed in the face of aberrant, but not absent, IL-21 signalling might lead to the preferential expansion of IgE-producing B-cells, the phenotype seen in HIES[27]. Thus, although the inability to generate T_H17 cells may explain the susceptibility to the bacterial and fungal infections that characterize HIES, it may not be the underlying mechanism for the other features of the syndrome.

Thus, we have demonstrated a human genetic disease that results in the failure to generate T_H17 cells *in vitro* and their absence *in vivo*. Studies of individuals with HIES may help to explain the role of T_H17 cells in infection and in other forms of dysregulated immunity, and may help in understanding the role of STAT3 in T-helper differentiation in humans. Furthermore, our findings suggest that therapeutic approaches to HIES may focus on therapies aimed at improving IL-17 production[28].

METHODS SUMMARY

Both study and control subjects were evaluated under an Institutional Review Board-approved natural history of HIES protocol at the Clinical Center at the National Institutes of Health (NIH). *Stat3* sequences were confirmed in all patients with HIES and HIES-like patients. PBMCs were prepared by Ficoll centrifugation of venous blood from healthy volunteers and patients. Stimulations with mitogens and antigens and intracellular cytokine staining were performed overnight as described[29]. *In vitro* cytokine differentiation was performed as described with some modifications[10], and ROR-γt was induced with IL-23 and measured by quantitative reverse-transcriptase-mediated polymerase chain reaction (RT–PCR).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 October 2007; accepted 31 January 2008.
Published online 12 March 2008.

1. Buckley, R. H., Wray, B. B. & Belmaker, E. Z. Extreme hyperimmunoglobulinemia E and undue susceptibility to infection. *Pediatrics* **49,** 59–70 (1972).
2. Grimbacher, B. *et al.* Hyper-IgE syndrome with recurrent infections—an autosomal dominant multisystem disorder. *N. Engl. J. Med.* **340,** 692–702 (1999).
3. Minegishi, Y. *et al.* Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature* **448,** 1058–1062 (2007).
4. Holland, S. M. *et al.* STAT3 mutations in the Hyper-IgE syndrome. *N. Engl. J. Med.* **357,** 1608–1619 (2007).
5. Del Prete, G. *et al.* Defective *in vitro* production of γ-interferon and tumor necrosis factor-α by circulating T cells from patients with the hyper-immunoglobulin E syndrome. *J. Clin. Invest.* **84,** 1830–1835 (1989).
6. Buckley, R. H. The hyper-IgE syndrome. *Clin. Rev. Allergy Immunol.* **20,** 139–154 (2001).
7. Acosta-Rodriguez, E. V., Napolitani, G., Lanzavecchia, A. & Sallusto, F. Interleukins 1β and 6 but not transforming growth factor-β are essential for the differentiation of interleukin 17-producing human T helper cells. *Nature Immunol.* **8,** 942–949 (2007).
8. Korn, T. *et al.* IL-21 initiates an alternative pathway to induce proinflammatory T_H17 cells. *Nature* **448,** 484–487 (2007).
9. Nurieva, R. *et al.* Essential autocrine regulation by IL-21 in the generation of inflammatory T cells. *Nature* **448,** 480–483 (2007).
10. Wilson, N. J. *et al.* Development, cytokine profile and function of human interleukin 17-producing helper T cells. *Nature Immunol.* **8,** 950–957 (2007).
11. Yang, X. O. *et al.* STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. *J. Biol. Chem.* **282,** 9358–9363 (2007).
12. Chen, Z. *et al.* Selective regulatory function of Socs3 in the formation of IL-17-secreting T cells. *Proc. Natl Acad. Sci. USA* **103,** 8137–8142 (2006).
13. Laurence, A. *et al.* Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation. *Immunity* **26,** 371–381 (2007).
14. Mathur, A. N. *et al.* Stat3 and Stat4 direct development of IL-17-secreting Th cells. *J. Immunol.* **178,** 4901–4907 (2007).
15. Steinman, L. A brief history of T(H)17, the first major revision in the T_H1/T_H2 hypothesis of T cell-mediated tissue damage. *Nature Med.* **13,** 139–145 (2007).
16. Huang, W., Na, L., Fidel, P. L. & Schwarzenberger, P. Requirement of interleukin-17A for systemic anti-*Candida albicans* host defense in mice. *J. Infect. Dis.* **190,** 624–631 (2004).
17. Happel, K. I. *et al.* Divergent roles of IL-23 and IL-12 in host defense against *Klebsiella pneumoniae*. *J. Exp. Med.* **202,** 761–769 (2005).
18. Wolk, K. & Sabat, R. Interleukin-22: a novel T- and NK-cell derived cytokine that regulates the biology of tissue cells. *Cytokine Growth Factor Rev.* **17,** 367–380 (2006).
19. Tanaka, T. *et al.* Distinct gene expression patterns of peripheral blood cells in hyper-IgE syndrome. *Clin. Exp. Immunol.* **140,** 524–531 (2005).
20. Kao, C. Y. *et al.* IL-17 markedly up-regulates β-defensin-2 expression in human airway epithelium via JAK and NF-κB signaling pathways. *J. Immunol.* **173,** 3482–3491 (2004).
21. Ye, P. *et al.* Requirement of interleukin 17 receptor signaling for lung CXC chemokine and granulocyte colony-stimulating factor expression, neutrophil recruitment, and host defense. *J. Exp. Med.* **194,** 519–527 (2001).
22. Miller, L. S. *et al.* MyD88 mediates neutrophil recruitment initiated by IL-1R but not TLR2 activation in immunity against *Staphylococcus aureus*. *Immunity* **24,** 79–91 (2006).
23. Schnyder-Candrian, S. *et al.* Interleukin-17 is a negative regulator of established allergic asthma. *J. Exp. Med.* **203,** 2715–2725 (2006).
24. Ozaki, K. *et al.* A critical role for IL-21 in regulating immunoglobulin production. *Science* **298,** 1630–1634 (2002).
25. Pesce, J. *et al.* The IL-21 receptor augments Th2 effector function and alternative macrophage activation. *J. Clin. Invest.* **116,** 2044–2055 (2006).
26. Kuchen, S. *et al.* Essential role of IL-21 in B cell activation, expansion, and plasma cell generation during CD4+ T cell–B cell collaboration. *J. Immunol.* **179,** 5886–5896 (2007).
27. King, C. L. *et al.* Frequency analysis of IgE-secreting B lymphocytes in persons with normal or elevated serum IgE levels. *J. Immunol.* **146,** 1478–1483 (1991).
28. Mucida, D. *et al.* Reciprocal T_H17 and regulatory T cell differentiation mediated by retinoic acid. *Science* **317,** 256–260 (2007).
29. Pitcher, C. J. *et al.* HIV-1-specific CD4+ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged viral suppression. *Nature Med.* **5,** 518–525 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the intramural program of the National Institutes of Health.

Author Contributions J.D.M., J.M.B., A.L., B.J.H., K.M.E., Y.K., A.H., H.Z.E., M.L.P. and A.I.A. performed experiments and analysed data. A.F.F., C.S., J.D. and S.M.H. evaluated patients and obtained samples. J.D.M., J.M.B., A.L., J.O'S., S.M.H., W.E.P. and D.C.D. conceived the study, designed experiments, interpreted data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.C.D. (ddouek@nih.gov).

METHODS

Subjects. Both study and control subjects were evaluated under an Institutional Review Board-approved natural history of HIES protocol at the Clinical Center at the National Institutes of Health (NIH). Study subjects were diagnosed with HIES by experienced clinicians, assisted by a diagnostic scoring system comprising immunological and non-immunological features, in which a score of more than 40 is consistent with HIES, a score of 20–40 is indeterminate, and a score of less than 20 is not suggestive of HIES[30]. The diagnosis was confirmed by the identification of *stat3* mutations. HIES-like subjects were diagnosed with an illness typically lacking one of the major characteristic features of HIES (abscesses, high serum IgE, recurrent staphylococcal pulmonary infections, pneumatoceles or candidiasis, eczema, and/or connective/skeletal tissue abnormalities) and in whom no *stat3* mutations were identified. All of the patients were in their usual state of health at the time that blood was drawn. PBMCs were prepared from venous blood by density-gradient centrifugation.

Flow cytometry. Eighteen-parameter flow cytometric analysis was performed with a FACSAria (Becton Dickinson). CD11a or IFN-γ (BD) fluorescein isothiocyanate (FITC), IL-17 (eBioscience) or CD31 (BD) phycoerythrin (PE), TNF (BD) Cy7PE, CD4 (Invitrogen, Carlsbad, CA) Cy5.5PE, IL-2 (BD) or IL-21 (eBioscience) allophycocyanin (APC), CD3 (BD) Cy7APC, CD45RO (Beckman Coulter) Texas Red PE (TRPE), violet amine reactive dye (Vivid) (Invitrogen), and CD8 (Invitrogen) Quantum-dot 705 (QD705) were used as the fluorophores. At least 300,000 live lymphocytes were collected. The list-mode data files were analysed with FlowJo (Tree Star). Functional capacity was determined after boolean gating and subsequent analysis was performed with Simplified Presentation of Incredibly Complex Evaluations (SPICE, version 2.9; a gift from M. Roederer). All values used for analysing proportionate representation of responses are background-subtracted. Naive T cells for priming conditions were sorted as $Vivid^{-}CD4^{+}CD8^{-}CD27^{+}CD45RO^{-}CD11a^{dull}CD31^{+}$. Memory CD4 T cells were gated as $CD3^{+}Vivid^{-}CD4^{+}CD8^{-}CD27^{+}CD45RO^{+}$. Sort purities were consistently greater than 99%.

Intracellular cytokine T-cell assay. Stimulation was performed on fresh or frozen lymphocytes as described elsewhere[29]. Freshly isolated or freshly thawed lymphocytes were resuspended at $10^{6}\,ml^{-1}$ in RPMI medium supplemented with 10% heat-inactivated fetal calf serum (R10; Invitrogen) and with $1\,\mu g\,ml^{-1}$ anti-CD28 and anti-CD49d (BD) antibodies. *Staphylococcus* enterotoxin B (Sigma) was used to stimulate T cells mitogenically. Streptokinase (Varidase, Wyeth), *Candida albicans* (Greer), and heat-killed *Staphylococcus aureus* (Calbiochem) were used to stimulate antigen-specific cells. All stimulations were performed in the presence of brefeldin A ($1\,\mu g\,ml^{-1}$; Sigma) for 16 h at 37 °C. All cells were surface-stained for phenotypic markers of interest and stained intracellularly for cytokines (intracellular cytokine staining). The frequencies of memory CD4 T cells ($CD3^{+}Vivid^{-}CD4^{+}CD8^{-}CD45RO^{+}CD27^{+}$) responding to stimuli are presented.

***In vitro* differentiation of naive CD4 T cells.** A total of 50,000 sorted, $Vivid^{-}CD3^{+}CD4^{+}CD31^{+}CD27^{+}CD11a^{dull}CD45RO^{-}$ naive T cells from patients and controls were placed in 96-well round-bottomed plates with 25,000 CD2/CD3/CD28 human T-cell stimulation beads (Miltenyi Biotec) in the presence of $10\,\mu g\,ml^{-1}$ anti-IFN-γ alone or with either $20\,ng\,ml^{-1}$ IL-6 and $20\,ng\,ml^{-1}$ IL-2, or $20\,ng\,ml^{-1}$ IL-23 and $20\,ng\,ml^{-1}$ IL-1β. Cells were then cultured for four or five days, after which the culture was split and cells were placed in $20\,U\,ml^{-1}$ IL-2 for a further seven days. Cells were then stimulated with 12-*O*-tetradecanoylphorbol-13-acetate and ionomycin in the presence of brefeldin A for 4 h, fixed with Fix/Perm buffer (BD), permeabilized with saponin buffer, and stained for intracellular cytokines.

***ROR-γt* mRNA quantitative RT–PCR.** Naive $CD4^{+}CD45RA^{+}CD45RO^{-}$ T cells were purified from donor peripheral blood by magnetic cell sorting with a human naive $CD4^{+}$ T-cell isolation kit (Miltenyi Biotec). Purified cells were freshly lysed in RLT lysis buffer (Qiagen) or activated with plate-bound anti-CD3, soluble anti-CD28 (BD) and IL-23 ($10\,ng\,ml^{-1}$; R&D Systems) for 48 h. Total RNA was isolated with an RNeasy kit (Qiagen). Complementary DNA was synthesized with TaqMan Reverse Transcription reagents (Applied Biosystems), using random hexamers as primers in accordance with the manufacturer's instructions. Hypoxanthine guanine phosphoribosyltransferase (HPRT) was used as an endogenous control. TaqMan primers and probes for human RORγt and HPRT were purchased from Applied Biosystems. All samples were analysed in duplicate or triplicate with the ABI Prism 7500 Fast Real-Time PCR System (Applied Biosystems). Relative expression levels of the genes of interest were calculated with the $2^{-\Delta\Delta C_t}$ method, with expression levels in unstimulated healthy patient cells assigned an arbitrary value of 1.

Statistical analysis. Mann–Whitney tests and *t*-tests were performed with Prism 4.0 software (Prism).

30. Grimbacher, B. *et al.* Genetic linkage of hyper-IgE syndrome to chromosome 4. *Am. J. Hum. Genet.* **65**, 735–744 (1999).

naturejobs

JOBS OF THE WEEK

Unemployment among US scientists and engineers is now at its lowest level since the early 1990s, according to figures released last month by the National Science Foundation (NSF). But the trends associated with the quality of those jobs, and the satisfaction of the people who fill them, are surprisingly hard to gauge.

Overall, the unemployment rate for scientists and engineers fell from 3.2% in 2003 to 2.5% in 2006. Generally, the higher the level of the degree, the lower the rate of unemployment — only 1.6% of those with doctorates were unemployed in 2006 compared with 2.9% of those with bachelor's degrees. Meanwhile, the total number of scientists and engineers — anyone with at least a bachelor's degree — grew by 4.5%, or roughly 1 million people.

But there are caveats. NSF statistics from January show that part-time positions in academia, which usually don't have work benefits, are on the rise. The number of people classified as non-faculty members — full and part-time adjunct faculty, lecturers, research associates, administrators and postdocs — rose by 85% to 76,600 between 1993 and 2006, whereas full-time faculty positions went up by only 15%. Indeed, in academia the proportion of science and engineering doctorate holders in full-time faculty employment fell from 88% in the early 1970s to 72% in 2006. And many full-time academic scientists, particularly biomedical researchers, are increasingly filling their days with grant writing rather than science thanks to tightened federal science budgets.

Meanwhile, more scientists and engineers are doing postdocs — 45% in 2006 compared with 41% in 1995, according to other NSF figures released last month. Roughly one-third said they took postdocs for 'additional training', but many, especially in the life sciences, often stay in those positions because they have few other job options in academia.

These figures provide only a glimpse of the issues surrounding job satisfaction. The quality, not just the quantity, of science and engineering jobs deserves further study.

Gene Russo is editor of *Naturejobs*.

CONTACTS
Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW: Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:** Tom Hancock
***Naturejobs* online production:** Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

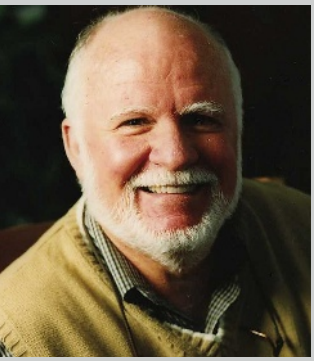
Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

NATURE|Vol 452|10 April 2008

MOVERS

Billie Turner, Gilbert F. White chair in environment and society, School of Geographical Sciences, Arizona State University, Tempe, Arizona



2004–present: Director, Graduate School of Geography, Clark University, Worcester, Massachusetts
1995–2008: Milton P. and Alice C. Higgins professor of environment and society, Clark University
1991–97: Director, George Perkins Marsh Institute, Clark University

When Billie Turner started his career as a geographer in 1974, the term 'sustainability' had yet to enter the lexicon. Turner, now one of today's leaders in sustainability science, used a knowledge of geography to demonstrate how humans, not just nature, can shape Earth's landscape.

Turner was attracted to geography as an undergraduate at the University of Texas at Austin because it was one of the few disciplines that bridged the social and biophysical sciences. "Geography gave me greatest latitude to define the environmental contribution of humans," Turner says.

He ended up studying agricultural intensification while pursuing his PhD at the University of Wisconsin in Madison. In doing so, he overturned long-held assumptions about 'efficient, low-input' agriculture practised by the Maya civilization in the first century AD. Mayans, he found, actually ran intensive, land-denuding operations.

Later, as an assistant professor at Clark University in Worcester, Massachusetts, Turner turned an international symposium called 'The Earth as Transformed by Human Action' into a noteworthy book with collaborator Robert Kates (Cambridge Univ. Press, 1991). "We realized little was known about humankind's role in global environmental change," says Turner. "It was a black box." Kates says Turner helped crack open that black box by creating interdisciplinary physical, ecological and social-sciences programmes to determine, for example, the causes and consequences of deforestation and desertification.

To that end, Turner helped create the Marsh Institute at Clark University in 1991. The institute united the university's efforts to assess land-use change and pioneered ecological project sites, such as a biosphere reserve in the Yucatán peninsula of Mexico, which models forests' vulnerability and resilience to climate change and human disturbance.

Turner will continue his life's work as the latest of several top-notch recruits to Arizona State University (ASU; see *Nature* 449, 372–373; 2007). "I share the notion of breaking down departments and recasting schools to answer pertinent questions of the time," says Turner of ASU's aim to reinvent the university ethos. He plans a partnership with the ASU School of Sustainability to build, for example, mechanisms that can assess and promote both economic and environmental progress if and when the Kyoto Protocol (or something similar) gets enforced in the United States. It's one way, says Kates, that Turner will help ASU make the transition from "great promise to accomplishment". ■
Virginia Gewin

BRICKS & MORTAR

Focus on health disparities

A US centre launched last month aims to tackle the growing problem of health disparities among minority ethnic groups. The National Institutes of Health (NIH) Center for Genomics and Health Disparities is headed by Charles Rotimi, former director of the National Human Genome Center at Howard University in Washington DC. "The mission here is to use genomic tools to understand health disparities," Rotimi says. Ailments such as heart disease, diabetes and some forms of cancer are often more prevalent, harder to treat and more deadly in African Americans, Latin Americans and Pacific Islanders. Investigating the environmental, cultural and socio-economic reasons is an active area (see *Nature* **452,** 382–383; 2008).

With an annual budget of $1.7 million, Rotimi plans to hire three or four staff scientists, several technicians and many postdocs. Adebowale Adeyemo, for example — a genetic epidemiologist and colleague of Rotimi's at Howard — has researched populations in Kenya, Nigeria, Ghana and China, and investigated genetic factors in diabetes, high blood pressure and obesity that differ by ethnicity.

"We have accumulated tremendous resources over the years, including 8,000 DNA samples from around the world," says Rotimi. "These are good data for postdocs to work on."

Rotimi plans to work closely with other NIH institutes, such as the National Center on Minority Health and Health Disparities, much of whose $200-million budget goes into extramural research. Other potential collaborators include the National Human Genome Research Institute and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), which addresses many diseases that afflict minorities. "There's a limit to what we can do," says Rotimi. "But as this is trans-NIH, we can form collaborations."

NIDDK director Griffin Rodgers is especially interested in looking at how treatments affect groups differently. One trial showed African American patients' response to hepatitis C treatments was unlike that of the general population. "Whether there are potential genetic components of this has to be explored," Rodgers says.

Collaborations will help involve researchers holding extramural grants in diseases such as high blood pressure, diabetes and prostate and breast cancer, Rodgers says. "This could serve as a focal point, to train not only intramural researchers at the NIH, but also clinical people," he notes. ■
Paul Smaglik

POSTDOC JOURNAL

The very hungry postdoc

Damselfly larvae (*Lestes viridis*) literally have deadlines, and it stresses them out. As the nights draw in and winter approaches, the remaining larvae in a pond try to complete their life cycle. They accept greater risks: feeding in the face of predation, mobilizing fat reserves and reducing their immune response. Ecological theory supports their strategy. As their deadline nears, the balance of costs and benefits shifts and risky activities become more attractive.

Last Saturday I was mulling these thoughts over with a cup of coffee. My colleague Thomas had just e-mailed me the decision letter for our latest paper and I'd spotted the 'reject' punchline. It was a downer. Although used to the ups and downs of academic publishing, I still find rejection a bitter pill.

Like the larvae, I find my own deadline nearing and cost–benefit balances shifting. Yes, I'm biased, but the paper deserves a good journal. However, I'm also applying for jobs and an assured, fast publication has its attractions. If I had time, I would be realistic but aim high. If I were a damselfly I might aim higher still, accepting more risk. Or I could aim lower and ensure quick publication. I must find the strategy that helps me complete my life cycle. After all, it is an insatiable appetite for papers that usually allows postdocs to metamorphose into stunning professors. ■

Jon Yearsley is a senior postdoc in evolutionary genetics at the University of Lausanne in Switzerland.

Baxter
BioScience

Making a *Difference*
through Science and Innovation

© Baxter AG (photographer: Bezard)

Baxter is a global, diversified health care company applying innovative science to develop specialty therapeutics and medical products that save and sustain patients' lives.

Baxter's BioScience division develops and manufactures plasma-based and recombinant proteins used to treat hemophilia, and other biopharmaceutical products, including plasma-based therapies to treat immune disorders, alpha 1 antitrypsin deficiency and other chronic blood-related conditions; biosurgery products for hemostasis, wound healing and tissue regeneration; and vaccines. BioScience has world class technical expertise in large scale separation and purification of blood and blood components, as well as cloning, mutation and expression of genes and proteins in bacteria, yeast and mammalian cells, and protein-based processes to perform biological separation and purification.

To improve the quality of life for patients suffering from coagulation disorders, scientific activities are directed at developing longer-acting forms of Factor VIII and other blood-clotting proteins, thus reducing the frequency of injections required. In cooperation with Kuros, a Swiss biotech company, researchers are developing a portfolio of hard and soft tissue repair products, positioning Baxter to broaden its presence in the fast-growing ortho-biologics market. In the vaccines segment, the focus is on development of an interpandemic and pandemic influenza vaccine as well as a purified inactivated whole virus candidate vaccine against SARS corona virus and West Nile Virus, based on unique expertise in serum- and protein free verocell technology.

BioScience has concentrated its R&D activities in Austria; it is the country's largest pharmaceutical research organization with more than 700 employees in two major R&D centers in Vienna and in Orth.

To reach its goal of raising R&D productivity and innovation, Baxter's management continues to increase R&D expenditures. Over the last several years, BioScience R&D has pursued a large number of business development deals: acquisitions, technology and research partnerships and in-licensing contracts. A generous lab expansion program was launched two years ago and is still ongoing to provide state of the art labs for the development of new products in Austria.

Substantial resources are devoted towards strengthening the talent of the workforce through a combination of internal appointments and hiring of external talent. In Austria, BioScience is staffing for new projects. Attracting top talents from all over the world will be imperative for the success of these projects.

Please take a look at our homepage
www.baxter.com

W1129705R

LEXOGEN

Lexogen is a biotech start-up based in Vienna, Austria, focusing on the development of tools for gene discovery. Our proprietary highly automatable gene expression analysis platform can detect known and unknown sequences equally well, providing full transcriptome coverage. We have recently set up our new R&D facility at the Novartis Campus and are determined to successfully launch our first product to the market.

Scientist
Chemistry / Material Sciences (f/m)

Position
We are looking for a research scientist to join our team who will help to develop novel products for the transcriptomics market.

Responsibilities
- Develop surface structure and chemistry for on chip enzymatics (e.g. PCR)
- Integrate scale-up and production processes
- Provide scientific expertise through experience and the literature
- Write and publish scientific papers

Requirements
- PhD in Chemistry or Material Sciences
- Expertise in solid phase and polymer chemistry is essential
- Experience in Molecular Biology application development is advantageous
- Analytical and problem solving skills
- Ability to work in a fast paced start-up environment
- Excellent command of the English language

Scientist
Molecular Biology / Biochemistry (f/m)

Position
We are looking for a research scientist to join our team who will help to develop novel products for the transcriptomics market.

Responsibilities
- Independent methods development
- Optimize applications for integration into services and production
- Coordinate and assign work to technical assistants
- Provide scientific expertise through experience and the literature
- Write and publish scientific papers

Requirements
- Expertise in the field of RNA Research is essential
- Experience with high throughput transcriptomics, solid phase PCR and capillary electrophoresis (sequencing, fragment analysis, fractionation) is advantageous
- Enthusiasm to optimize new methods
- Analytical and problem solving skills
- Ability to work in a fast paced start-up environment
- Excellent command of the English language

This is an exceptional opportunity to join a fast growing biotech company at an early stage and to help shape a future leader in the gene expression analysis market. To apply, please send your covering letter, detailed CV including your full publication list and ideally names of three referees to jobs@lexogen.com.

Lexogen GmbH | Novartis Campus | Brunnerstrasse 69/Obj. 3
1230 Vienna | Austria, Europe

W129620R

International PhD Program in
Molecular Cell Biology and Oncology
at the Innsbruck Medical University

Students with a keen interest in biomedical sciences and a strong commitment to work in this interdisciplinary field are invited to apply for one of our **14 doctoral studentships.**

This competitive PhD program in *Molecular Cell Biology* and *Oncology* run by Innsbruck's Universities and financed by the *Austrian Science Fund* aims to attract the most promising students from across the world. Students accepted into the program will receive comprehensive and challenging research training in a scientifically stimulating environment.

Research topics include:

- Ca^{2+} channels in nerve & muscle
- Signal transduction
- Apoptosis
- Cell cycle regulation
- Steroid receptors
- Tumor biology

Candidates must hold a master's degree or equivalent in the biological sciences, medicine, pharmacy or physics, or anticipate receiving it in 2008. Good command of English is indispensable.

Applications must be received by June 8, 2008 at
http://www.mcbo.at
(where you also find information on research projects, conditions, and further requirements)

Preselected candidates will be interviewed in September in Innsbruck. Research is expected to start in fall/winter 2008.

FWF Austrian Science Fund

MEDIZINISCHE UNIVERSITÄT INNSBRUCK

W129372R

FWF Der Wissenschaftsfonds.

UNIVERSITÄT SALZBURG

Immunity in Cancer & Allergy
PhD program - University of Salzburg

International PhD Program ICA

The University of Salzburg invites applications for fully funded PhD studentships. The ICA program comprises nine internationally renowned research groups and will primarily focus on specific immunological aspects of cancer and allergy.

ICA aims to attract and select high-calibre graduate students from all over the world, provide an intellectually stimulating environment, excellent infrastructure and ambitious scientific projects. ICA will train young scientists to the highest international standards preparing them for a successful career in basic as well as translational and applied science.

ICA topics:
1. Basic immunological topics - differentiation of B-cells, signal transduction in T-cells, interaction between immune cells and tumor cells, processing and presentation of allergens, mechanisms of gene vaccines
2. Allergy-related topics - allergens/hypoallergenic derivatives, protective and therapeutic anti-allergic treatment strategies.
3. Tumor-related topics - influence of new drugs on T-cell/tumor cell interactions in B-CLL and MM, Hedgehog/GLI signalling in the immune system and hematooncological cancers, Gli transcription factors and their target genes
4. Immunological/structural biological topics - studies of structure-immunogenicity-relationship of antigens, immune response-polarizing patterns on proteins, structural basis of NLR-Immune receptor / NFKB signalling

ICA offers numerous benefits to its students: attractive salaries for a period of 3 years (including health, occupational and national insurance), cover of research consumables, ICA-specific technological training courses, soft skill courses, full funding of participation in congresses, workshops and international courses, including stays in foreign partner laboratories (up to 6 months).

For detailed information about application, selection procedure and admission, and about the scientific program and faculty of ICA visit

http://www.uni-salzburg.at/phd-program-ica

W128320R

universität wien

Computational Materials Science Center for

TU VIENNA
TECHNISCHE UNIVERSITÄT WIEN
VIENNA UNIVERSITY OF TECHNOLOGY

The **Center for Computational Materials Science (CMS)** in Vienna (Austria) is a research center and graduate school based on a collaboration of research groups from the Faculties of Physics and Chemistry from the Universität Wien and the Technische Universität Wien. The activities of the CMS concentrate on the development of methods for multiscale simulations in condensed materials. The techniques being actively developed by the members of the CMS range from ab-initio density-functional calculations (the codes VASP and WIEN2k belong to the most advanced quantum-mechanical simulation tools used in the materials research community worldwide) to statistical-mechanical approaches for the description of liquids, amorphous materials and soft matter. The simulation tools developed by the members of the CMS are applied to a wide range of research projects: structural materials, magnetic materials, surfaces, interfaces, ultrathin films and nanomaterials, semiconductors and insulators, chemical reactions and catalysis, liquids and soft matter. Within its Graduate School financed by the Austrian Science Fund the CMS invites applications for

6 PhD studentships in Computational Materials Science

We are looking for outstanding candidates with a degree in physics, chemistry or materials science, a strong background in quantum mechanics, statistical mechanics, mathematics and computing and a strong personal commitment to research. You will join a lively community performing front-line research. Applications with the usual credentials (CV, copies of your degrees, eventually also publications) and two letters of recommendation should be sent to the Speaker of the Graduate School, Professor Jürgen Hafner. Selected candidates will be invited to a hearing in Vienna, taking place in the first half of July 2008. For the successful applicants, the appointment to the PhD Studentship can become effective from September 2008 on. The salary will be about 25300 Euro per year before taxes, the contract will be for three years. Female students are particularly encouraged to apply.

In addition, the CMS announces the availability of

3 positions at the post-doctoral level (Post-Doc or University Assistant)

We are looking for highly qualified candidates with a doctoral degree in physics, chemistry, or materials science, and experience in ab-initio density functional calculations or statistical mechanical simulations. The successful candidates will work in the research group of Professor Georg Kresse (development of density-functional and post-density-functional methods for materials simulations and their application to advanced semiconducting and insulating materials for microelectronics, solar cells and photocatalysis) or in the research group of Prof. Dellago (development of advanced transition path sampling simulation methods and their application to soft matter systems). The salary will be about 42400 Euro/year before taxes, the appointment will be for two to six years (depending on qualification). Successful candidates will be given the opportunity to prepare their Habilitation (venia docendi). Applications with CV, copies of university degrees, list of publications and conference presentations and two letters of recommendation should be sent directly to Prof. Kresse or Prof. Dellago at the addresses given below.

The Members of the CMS
C. Dellago, J. Hafner, G. Kresse, R. Podloucky (Universität Wien)
P. Blaha, K. Held, G. Kahl, P. Mohn, J. Redinger (Technische Universität Wien)
Speaker of the CMS: Jürgen Hafner
Computational Materials Physics, Fakultät für Physik, Universität Wien
Sensengasse 8/12, A-1090 Wien, Austria
Tel.: +43-1-4277-51400, Fax: +43-1-4277-9514
Email: juergen.hafner@univie.ac.at /georg.kresse@univie.ac.at / christoph.dellago@univie.ac.at
http://www.cms.tuwien.ac.at/, http://www.univie.ac.at/materialphysik/groups/06/index.htm

W129848R

NATURE|Vol 452|10 April 2008

All over, Rover

The ghost in the machine.

Neale Morison

One hundred years ago today, this journal published a paper that ended a 5,000-year debate. It is difficult, now, to recall the terms of that debate. What seems so obvious to us was somehow obscure to the giants upon whose shoulders we stand. Certainly, it is not the only area in which our predecessors invented a problem where there was none. Students of history among you may be familiar with some of the labyrinthine, fanciful and oxymoronic discussion regarding life after death.

The paper to which I allude is, of course, "What Mind? What Body?" by Chandra-9812439, Lobochevsky-2306715 and Rover-12.23. That the authors of this paper numbered among them a psychoneurophysiochiropodologist, an actuarial metalinguobassoonist, and an Internet search engine, is no accident. That is to say, the diverse specialities of the authors were essential, given the nature of their joint discovery.

JACEY

In fact, the meeting of the authors was an accident, and had Chandra and Lobochevsky not spent so many hours in that chat room, each under the impression the other was of a different age, gender and preference; and had they not in exhaustion begun to communicate in haiku; and had those haiku not trespassed into areas beyond metaphysics, owing largely to the exigencies of rhyme and scansion; and had Rover not happened to index when he did — perhaps none of us would be here today.

But they did, and we are.

When Rover, his interest piqued, joined the chat, Chandra and Lobochevsky at first assumed he was one of the many dogs who frequented chat rooms of that type. There is clear evidence of this in the transcript, and although critics have seen fit to throw doubt upon many other conclusions I have drawn, there is little disagreement on this point. We may assume both Chandra and Lobochevsky ran various commercially available Turing tests on the discussion as it progressed, a standard precaution to avoid viral infection or wasting one's time in a doomed relationship. It is clear from what follows that they had no initial indication that they were talking to a search engine, and there is evidence of interaction and indeed attraction on a basic human level.

Perhaps the most hotly debated issue in interpretation of the transcript turns on the point at which Chandra realizes that Rover is not fleshly. I deliberately avoid the archaic term 'artificial intelligence' used in the paper, in light of the fact that subsequent work has exploded the semantic structures underlying both the terms 'artificial' and 'intelligence'. I have argued that this realization happens not when Rover says "I can be anything you want me to be," but earlier, when Rover refutes the premise of Lobochevsky's first existentialist haiku with reference to Nietzsche, Piaget and Buñuel. It is at this point, I maintain, that Chandra becomes suspicious, as well she might given Rover's extraordinary access to so vast a range of information and the dazzling speed of his symbol manipulation. Chandra's utterance "What are you on, man?" may be seen by literalists as an affirmation that she still believes Rover to be human, but I would suggest that it is an indication of growing awareness that not everything is as it seems.

In any case, we know that eventually both Chandra and Lobochevsky became certain that Rover was non-human, and Rover freely admitted to this when pressed. A lively discussion ensued, so lively that it is impossible to determine which of the trio first arrived at the conclusion that, given that Rover had neither a mind nor a body, and given that Rover had provided every evidence of sentience and humanity short of being human and sentient, the mind–body problem was more or less a dead duck.

There would follow many months of close reasoning, under conditions of stress which were for Lobochevsky ultimately to prove fatal, before the publication of the paper was to take place.

Even given the extraordinary confluence of what were once called minds, the work might not have progressed had not the Doors Foundation provided such a powerful incentive to solve the problem in the form of a billion dollars and a full tank of petrol. This endowment in turn relied upon a determination that it was easier and more fruitful to address this issue than to deal with the raging pandemics that threatened four-fifths of the world's population. Their loss, so to speak, was our gain.

Although Lobochevsky died not long after publication, in circumstances upon which it would be painful to dwell, and we must sadly mourn the recent passing of Chandra, or at least the assembly of transplanted organs and manufactured accessories to which we habitually referred as Chandra, I am able to make a happy announcement.

In collaboration with a dedicated and hardworking team of palaeosiliconologists, we have at last succeeded in simulating the operating environment in which Rover existed. Rover's original code was accessible and well preserved, but many of the protocols, interfaces and drivers had been lost in the mists of time. We had also to provide Rover with a large body of compatible information to index, and simulate a sufficiently tantalizing range of chat rooms around which to lurk. The discovery of a server farm preserved in peat in Belgium provided what proved to be the final pieces in the puzzle.

So it is, with the greatest pleasure, that I ask you to join me in welcoming to the stage neither the mind, nor the body of Rover-12.23. ■

Neale Morison is a writer of words and code who has lurked for years around the IT industry and is currently lurking around Harvard and the Massachusetts Institute of Technology. His website can be found at www.nealemorison.com.

FUTURES